

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 12:21:43 ; Search time 4228 seconds
(without alignments)
16632.840 Million cell updates/sec

Title: US-10-082-894-1
Perfect score: 1719
Sequence: 1 gtttaattaccgaagtgtga.....aaatctccgtataaaaatt 1719

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_em.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_un.*
28: em_vi.*
29: em_htg_hum.*
30: em_htg_inv.*
31: em_htg_other.*
32: em_htg_mus.*
33: em_htg_pln.*
34: em_htg_rnd.*
35: em_htg_mam.*
36: em_htg_vrt.*
37: em_sy.*
38: em_htgo_hum.*
39: em_htgo_mus.*
40: em_htgo_other.*
41: em_htgo_mus.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	360.8	21.0	1563	8	AB033336	AB033336 Aspergill
C 2	336.2	19.6	299850	1	AF004601	AF004601 Oceanobac
C 3	321.8	18.7	36296	3	AF039713	AF039713 Caenorhab
C 4	314.2	18.3	248050	1	AL596172	AL596172 Listeria
C 5	314.2	18.3	349980	6	AX413018	AX413018 Sequence
C 6	314.2	18.3	349980	6	AX417048	AX417048 Sequence
C 7	312.6	18.2	324050	1	AL591983	AL591983 Listeria
C 8	312.6	18.2	349980	6	AX641672	AX641672 Sequence
C 9	301.2	17.5	301150	1	AF003190	AF003190 Clostridi
C 10	299.6	17.4	1654	1	AF120090	AF120090 Bacillus
C 11	299	17.4	300854	1	AE017014	AE017014 Bacillus
C 12	292.6	17.0	291205	1	AE017040	AE017040 Bacillus
C 13	291.4	17.0	5111	8	ASPLAS	X64705 Antithamio
C 14	287.2	16.7	12062	1	AE007586	AE007586 Clostridi
C 15	284.8	16.6	12844	1	U39725	U39725 Mycoplasma
C 16	284.8	16.6	80073	6	AR300198	Continuation (6 of
C 17	283.8	16.5	303650	1	AP001519	AP001519 Bacillus
C 18	282.6	16.4	300050	1	AP004171	AP004171 Mycoplasma
C 19	272.4	15.8	296650	1	AP005082	AP005082 Vibrio pa
C 20	271.2	15.8	300242	1	AE016745	AE016745 Staphyloc
C 21	271	15.8	303671	1	AE015937	AE015937 Clostridi
C 22	269.6	15.7	3373	1	AF269838	AF269838 Staphyloc
C 23	269.6	15.7	3373	6	AX145156	AX145156 Sequence
C 24	269.2	15.7	191028	8	PPU38804	U38804 Porphyra pu
C 25	266.2	15.5	1518	6	AX143497	AX143497 Sequence
C 26	266	15.5	300169	1	AE016801	AE016801 Vibrio vu
C 27	261.8	15.2	37669	1	AC074317	AC074317 Staphyloc
C 28	260.2	15.1	299050	1	AP003131	AP003131 Staphyloc
C 29	260.2	15.1	348527	1	AP003360	AP003360 Staphyloc
C 30	258.6	15.0	290150	1	AP004824	AP004824 Staphyloc
C 31	258.2	15.0	1515	6	AX619984	AX619984 Sequence
C 32	255	14.8	11066	1	AE013608	AE013608 Versinia
C 33	255	14.8	288050	1	AJ414141	AJ414141 Versinia
C 34	253.6	14.8	209510	1	BS000018	299121 Bacillus su
C 35	253.4	14.7	301903	1	AE016969	AE016969 Mycoplasma
C 36	247.6	14.4	10029	1	AE010963	AE010963 Methanosa
C 37	247.2	14.4	3895	1	BACTPMPEN	L29475 Bacillus su
C 38	247.2	14.4	347550	1	AP003595	AP003595 Nostoc sp
C 39	245.4	14.3	10322	1	AE013316	AE013316 Methanosa
C 40	242.8	14.1	10701	1	AE000021	AE000021 Mycoplasma
C 41	240.2	14.0	301660	1	AE016768	AE016768 Escherich
C 42	239.4	13.9	304230	1	AE016940	AE016940 Bacteroid
C 43	238.2	13.9	245050	1	AL627280	AL627280 Salmonell
C 44	238.2	13.9	300102	1	AE016847	AE016847 Salmonell
C 45	235.6	13.7	1586	1	AF120091	AF120091 Bacillus

ALIGNMENTS

RESULT 1
AB033336
LOCUS
DEFINITION Aspergillus oryzae gpmA mRNA for phosphoglyceromutase, complete cds.
ACCESSION AB033336
VERSION AB033336.1 GI:9955874
KEYWORDS Phosphoglyceromutase.
SOURCE Aspergillus oryzae
ORGANISM Aspergillus oryzae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 1563)
AUTHORS Nakajima,K., Kunihiro,S., Sano,M., Eto,S. and Machida,M.
TITLE Molecular cloning and characterization of glycolytic gene from

TITLE	Genome sequence of Oceanobacillus iheyensis isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments
JOURNAL	Nucleic Acids Res. 30 (18), 3927-3935 (2002)
MEDLINE	22220767
PUBMED	12235376
REFERENCE	3 (bases 1 to 299850)
AUTHORS	Takami,H., Takaki,Y. and Chee,G.
TITLE	Direct Submission
JOURNAL	Submitted (26-DEC-2001) Hideto Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group; 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan
	(E-mail:takami@jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html, URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html, Tel:81-468-67-9643, Fax:81-468-67-9645)
FEATURES	Location/Qualifiers
source	1..299850
	/organism="Oceanobacillus iheyensis HTE831"
	/mol_type="genomic DNA"
	/strain="HTE831"
	/db_xref="taxon:221109"
	/note="halotolerant and alkaliphilic"
gene	95..487
CDS	/gene="OB2343"
	95..487
	/gene="OB2343"
	/note="CDS_ID OB2343"
	/codon_start=1
	/transl_table=11
	/product="hypothetical conserved protein"
	/protein_id="BAC14299.1"
	/db_xref="GI:22778029"
	/translation="MKTISPRCSLITSATPLRASSSFPSAETISAPRGIPKDIPT MIDPKLCLPLEYTSISESNCPALFTKNVGRACPTVSAINVNCDEITHTPSPYEL YDCLCLSIITLNLISHRPPIIYKHKR"
gene	complement(106..369)
CDS	/gene="OB2344"
	complement(106..369)
	/gene="OB2344"
	/note="CDS_ID OB2344"
	/codon_start=1
	/transl_table=11
	/product="PTS system, histidine-containing phosphocarrier protein (HPr protein)"
	/protein_id="BAC14300.1"
	/db_xref="GI:22778030"
	/translation="MKSQTFITAEVGVARPATLLYNKAGQFSEIEVSYKGQVNL KSMGVSLGIPKGAIEVSDAGKDEEDALNGVAEVIKEHGE"
gene	590..1297
CDS	/gene="OB2345"
	590..1297
	/gene="OB2345"
	/EC_number="2.4.2.1"
	/note="CDS_ID OB2345"
	/codon_start=1
	/transl_table=11
	/product="purine nucleoside phosphorylase"
	/protein_id="BAC14301.1"
	/db_xref="GI:22778031"
	/translation="MSVHIGAKQGEIADKILLPGDPLRAKYAETFFLEEPKLYNEVRG MYGTGYKGERVSQGTGMQVPSISIVNELIQSDYVKLLRVGTCGAIOKQVNVRD VLAQAGATTOSMRLIFNGIDYAPIADFLLKNYDAGVDKGLHLRAGVNGVFTSDTFY RDNAAQVNEILLAKYQVLAIEWESTALYTLAAKYRQALSVLITVSDHITIGETTAER QTTENDMMETALDILIK"
gene	1422..2021
CDS	/gene="OB2346"
	1422..2021
	/gene="OB2346"
	/note="CDS_ID OB2346"
	/codon_start=1
	/transl_table=11
	/product="biotin biosynthesis"
	/protein_id="BAC14302.1"
	/translation="MQQFRAIHITYTSLFVCLMVIGANIAIWFPLFVPIGGSVPLS LOTFFAILAGMILKRGAFAMIVYILIGLTGLPVFAGLKGLYPMISPTGCFIISFV FIATIVGMLLENKSVSKDLITVAIIGLMINYTLGYVYMIANNWLELPISLITAW IGWVPFLVKDSVLITVAIITMMALQKRVSSIPVIONISR"
gene	2117..2596
CDS	/gene="OB2347"
	2117..2596
	/gene="OB2347"
	/note="CDS_ID OB2347"
	/codon_start=1
	/transl_table=11
	/product="hypothetical conserved protein"
	/protein_id="BAC14303.1"
	/db_xref="GI:22778033"
	/translation="MTLFSNFWALCAIYFAOVIKIPIKLIETKEPOPCGLAFSTGG MPSSHAAYTALTITTAIGITEGTVSSVFALACIFSVITMFEDASGVRRQAGEAYVINQL IRDFQLITTSKADWNKKGTEKQLQELKHGHQPIEVFFGALTGIIILAFSTAPLSY"
gene	complement(2702..3375)
CDS	/gene="OB2348"
	complement(2702..3375)
	/gene="OB2348"
	/note="CDS_ID OB2348"
	/codon_start=1
	/transl_table=11
	/product="hypothetical conserved protein"
	/protein_id="BAC14304.1"
	/db_xref="GI:22778034"
	/translation="MKIKNMKMHVGMLSIICLSVLETSVTNLFIPKRVKANPIELIP SSEGIIIEQFNKSLKPNELVAKANOTLLSNDIEGSSIDEADLEQFPVHTVMTG YTAGIESTGKTAHEPGYCTIFSGVQVRDLSTIADLSVYPITGILFIPDYGCVVA DKGSALTGNKIDLYHTVDVYEQMKKEVDVYVIEMGDGTLTBETLLELNEDESMQV FROQFN"
gene	complement(3438..3749)
CDS	/gene="OB2349"
	complement(3438..3749)
	/gene="OB2349"
	/note="CDS_ID OB2349"
	/codon_start=1
	/transl_table=11
	/product="hypothetical conserved protein"
	/protein_id="BAC14305.1"
	/db_xref="GI:22778035"
	/translation="MVOLIVSVFLYFVIFFGIAFLNMLRQTLWMAFYPIVVLII DNVCMAEYFTNTATAFNDLSTFLDLIAFDIVILAGFIGTIVSGFVTKLRKSYQM F"
gene	complement(3849..5060)
CDS	/gene="OB2350"
	complement(3849..5060)
	/gene="OB2350"
	/note="CDS_ID OB2350"
	/codon_start=1
	/transl_table=11
	/product="NADH dehydrogenase"
	/protein_id="BAC14306.1"
	/db_xref="GI:22778036"
	/translation="MKNKPHIVILGAGYGMMAARMOKLIHSNEAKITLVNKNDYHYO TTWLHENAAGTLHHRDSKIDIKVIDNRKILILDEVVSIKREKKYVKLKKQELYYDI LIVSLGEPALYIGPSEFAYSINDRSARLIERIEINFALTYENKPDGRNLIV IGGGLGVFEFAGELINRIPTICEYDIEKNLVKYLLEGRDCLVLYGYDSSAVQYAMN SLOQRGVEIINDATVKADESIIVKRGSEHKISTKSVIWTAGTKANSIVERLNLMS EHGVRKYRDLRSIEDDSIYLYIGDCSLNYDENGVLVPSAQIAMQOGEFIARNINAKI KGEKMAAFPLDPVPGRIITSLGSDDGIAOMFSRNLFGWKAALLKRVADNWYLFKLGGIQL LMKXGKNMFH"
gene	5424..6410
CDS	/gene="OB2351"
	5424..6410
	/gene="OB2351"
	/note="CDS_ID OB2351"
	/codon_start=1
	/transl_table=11
	/product="thioredoxin reductase"
	/protein_id="BAC14307.1"

```
/db_xref="GI:22778037"
/translation="MDSKVFDTIIGAGPTGLFTAFYGMREASVKIIESLPHIGSQL
TALYPEKYIYDAGFPKVAQELIDRLAEQAFDFDEIIVLGOAIDKVERLEDETRILT
SNTGEVHLTKTIIYAGNAGFQPRRLNIGNSEAFEGKNLHYVVRDMNOYKGRVYLLG
GDSADVHMLLEKIAKVTLYVRRDQFRAHESVNOLKESSEVEILLTPYAPINVEASD
KIEKIQLEKVEKIDIVEDDIIICNGYFISLGPDKMGLEIKRNSIIVNSKMETNI
PGVYAGDVCTYDGRVKLGTATGFGGPTAINNKNKNIIDPKARIQPKHSTAMF"
gene
CDS
complement(6465..7061)
/gene="OB2352"
complement(6465..7061)
/gene="OB2352"
/note="CDS_ID OB2352"
/codon_start=1
/transl_table=11
/product="chitooligosaccharide deacetylase (modulation
protein) (partial)"
/protein_id="BAC14308.1"
/db_xref="GI:22778038"
/translation="MFQORERMIALTFDDGPLPGPTNEILDILSENOSRATFVLGNN
VTKHRDVRRIVREGHDIQVHSFSPNFQETEIYEVRELIDSTLLELDLGLVLYV
RPPYGFSTAETQFLDQIGLSILTYDTLDWEILNAEOIASNLLENIQAGDIVLLHD
IHPTIOAMRIAIPKLVQGYKLVTISEWLEQENPFK"
gene
CDS
complement(7514..7870)
/gene="OB2353"
complement(7514..7870)
/note="CDS_ID OB2353"
/codon_start=1
/transl_table=11
/product="hypothetical conserved protein"
/protein_id="BAC14309.1"
/db_xref="GI:22778039"
/translation="MIINISASEQIKEMKDESDQSHLRFKIRGGGSGLSYSLGF
DEDINEELDIVIONINGIPVVFNFQDPILEGTTIDYKENLGGGFSIDNPNAIVSGC
GSSFKDKVGPQDC"
gene
complement(7943..8803)
Query Match 19.6%; Score 336.2; DB 1; Length 299850;
Best Local Similarity 54.3%; Pred. No. 4.5e-63;
Matches 752; Conservative 0; Mismatches 618; Indels 15; Gaps 3;
QY 65 TTATTCATGATGGGCGCTTCCGATGAACACACACGGGAATCAATGCTAAAGCTAA 124
Db 92365 TTCCTGATGATTCGCTATTCGTGATGAAGTAAAGGAATGCGATCAACAGCAACA 92306
QY 125 CGCTATTATGGCAAACTTTGCTCGAAATGGCAAA---AATTGGAAGCACACGGTC 181
Db 92305 CCCCTAAATTTGATCGTTTGGAAATCAATATGCTCATATCAATTAGAGGCTCTGGAA 92246
QY 182 TTCATGTTGGATGCCAGAGGCTTAATGGGAAATTCGAAGTTGGACATTTGAATATAG 241
Db 92245 AGGCTGTTGGATTACAGATGGCAATGGGAAATCCGAATTTGGTCATTTGAATATTG 92186
QY 242 GAGCTGGAAGACTTATTAATCAAGATATGTTGCAATTAATTTGGCTGTTCAACGAAACG 301
Db 92185 GTGCTGGTGGTATTTGATCAAGAGCTTACTCGTGAATATGTTGATGACGAGGTTG 92126
QY 302 AGTTTGTACAAATCCCTCAGATTTGTCATCAGCTGAGCGTCAAGAGAGGGAGTGGTC 361
Db 92125 ATTTCTATGAGTTGATGCGTTTCATTAATCTGTAGAAATGCAAGAATCATGACAAAG 92066
QY 362 GATTCGATTTATAGGACTGGTTAGCGATGGTGGTGTCCACTCTCATATGATCATCTTTT 421
Db 92065 CGCTTCACATTTATTTGTTTATATCTGACGGTGGTGTGATAGTATATTAATCATTTAT 92006
QY 422 TTGCGTTGATACGCTGATTTAAACATTAACATGACGCAAGGTTTTCATCTACTTTTG 481
Db 92005 TTGCTTTACTTCGATTAGCGAAGAACATGAATTAAGAAATGATTGTACATGCTTTCT 91946
QY 482 CTGATGGTGCAGATACCTCGCCCAACAAGTGGAGCTGGTTATCTTGAACAACATCTTCAAT 541
Db 91945 TAGACGCGCTGATGTTGGTCCGAAACAGCGCAAGCAATATATCAAGAAGAACACAGATG 91886
```

RESULT 3

AF039713

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

AF039713

Caenorhabditis elegans

36296 bp DNA linear INV 26-DEC-2002

complete sequence.

AF039713.1

HTG

Caenorhabditis elegans

ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 36296)
 AUTHORS Waterston,R.
 TITLE Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 JOURNAL 99069613
 MEDLINE 9851916
 PUBMED
 REFERENCE 2 (bases 1 to 36296)
 AUTHORS Greco,T., Elliott,G. and Keppler,D.
 TITLE The sequence of C. elegans cosmid F57B10
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 36296)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (23-DEC-1997) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 4 (bases 1 to 36296)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (19-JUL-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 5 (bases 1 to 36296)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (19-APR-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 6 (bases 1 to 36296)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 7 (bases 1 to 36296)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (23-OCT-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 8 (bases 1 to 36296)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (19-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 9 (bases 1 to 36296)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (26-DEC-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 COMMENT Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: rwnematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from

more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=F57B10;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is T10E9, 300 bp overlap; the 3' cosmid is F48A9,
 4100 bp overlap. Actual start of this cosmid is at base position
 197 of F57B10; actual end is at 6600 of F48A9.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data : computer analysis using the program
 Genefinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yuji Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFome cloning project (<http://wofdb.dfci.harvard.edu/>),
 similarity to other proteins from BlastX analyses
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual C. elegans GenBank submissions,
 and personal communications with C. elegans researchers. TRNAS
 are predicted using the program tRNAscan-SE (Lowe, T.M. and
 Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source

1. .36296
 /organism="Caenorhabditis elegans"
 /mol_type="genomic DNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="1"
 /clone="F57B10"
 /complement(417..1966)
 /gene="F57B10.8"
 complement(join(417..623,676..757,801..970,1018..1119,
 1648..1799,1873..1966))
 /gene="F57B10.8"
 /standard_name="F57B10.8"
 /note="contains similarity to Oryctolagus cuniculus
 TROPONIN T, CARDIAC MUSCLE ISOFORMS (TNTC).
 SW:TRT2.RABIT; coded for by the following C. elegans
 cDNAs: CEESL47F"
 /codon_start=1
 /product="Hypothetical protein F57B10.8"
 /protein_id="AAB96725.1"
 /db_xref="GI:2773208"
 /db_xref="WormBase:F57B10.8"
 /translation="MAPIPKKKIVRSIKKPVVEEDMKMEKEKMLEPKEESIETD
 NNEDEAGNLDSDSEIKTESEDEDETEDEQKQKKTGVVFFSIPKYNVVRREYF
 EKRCPGQVGRIFLARNKSNPQYLYSEGWMELKKIARAIAEQIDNTPIGGRKNDP
 VSSMLNRYLGSFKFVHLMQEQYKEVEKRMNVNVAQARRIAAHFEQIEKRGKL
 RKLKAVTAGSGKWKDFQVDVQKQKVVVKVKKERSKHVNTESSELMNNIFK"
 complement(2302..4228)
 /gene="F57B10.9"
 complement(join(2302..2636,2689..2796,2917..3285,
 3332..3536,3616..3711,3756..3837,3943..4067,4118..4228))
 /gene="F57B10.9"
 /standard_name="F57B10.9"
 /note="contains similarity to Interpro domain IPR000169
 (Eukaryotic thiol (cysteine) protease); coded for by the
 following C. elegans cDNAs: yk281e10.5"
 /codon_start=1
 /product="Hypothetical protein F57B10.9"
 /protein_id="AAB96726.2"
 /db_xref="GI:20198842"
 /db_xref="WormBase:F57B10.9"
 /translation="MFSIFNRIPDTAQVDVAFSCAYASIEGCLCYDEVNDWENTLAMY
 EKGUNLIVEGEKMKMKNEMKMLQESKSVQHRINVLRKRGPKRPPKRAEES
 VENKEAIRQLDLSFGQDLTYLPEGVQLFTIDGKTAPTAPTSLQTLRFPQT
 DGGASSDTLAFMQVGPWAYPLMGAKTPVLRNFGAYLVANPTPENNTVAILLSSDI

gene

CDS

```

ERRLIELHIVLREFTDFKEQAPSTELSKEENRISTTISNFIKGGOKIANGVETT
TVRISVEDNGQYRTTLTATQPMQSPVIGKSVYVYHKGKTKVAKCTRYLLDKIG
DMGVSKGKLIADSAORFEGDKSGSLVSGTIEILGGITCVATVMSLEDGSRHICRS
IANQTVQVNLKYGDDASDTHALFAAGHGTILAAQWLDLGPRSVAGRWARKAGIQM
VSDHHSRSGSLDQVEPFALEKKAV"
/genes="F57B10.14"
/complement(join(4552..4987))
/genes="F57B10.14"
/complement(join(4552..4666,4932..4987))
/standard_name="F57B10.14"
/notes="coded for by the following C. elegans cDNAs:
yk1069d02.5, yk1069d02.3, yk1107g08.3, yk1019c04.3"
/codon_start=1
/product="Hypothetical protein F57B10.14"
/protein_id="AA081096.1"
/db_xref="GI:21913056"
/db_xref="WormBase:F57B10.14"
/translation="MLYMKPAREVKNFTNPVFKSYGAYGGSFAFLIAYFCWEKTVG
QYIPLWNKRYVE"
/complement(5230..5905)
/genes="dad-1"
/complement(join(5230..5360,5695..5905))
/genes="dad-1"
/standard_name="F57B10.10"
/notes="contains similarity to Pfam domain PF02109 (DAD
family); coded for by the following C. elegans cDNAs:
CEESW25F, yk115g3.5, X89080"
/codon_start=1
/product="Dad (defender against apoptotic death) homolog
protein 1"
/protein_id="AAB96727.1"
/db_xref="GI:2773210"
/db_xref="WormBase:F57B10.10"
/translation="MAQVVPVLSKLEDDYOKTYSKKLIIDAYMYIILFTGIFQFIY
CLLVGFPPNSFUSGISVTSFVLSCLMNVQENRSEFTAVSTERAFADFIFANL
ILHLVVVNFILG"
/genes="F57B10.7"
/join(6517..6545,6591..6691,7177..7400,7446..7546,
7593..7685,8307..8720,8904..9207,9252..9542,9801..9947)
/genes="F57B10.7"
/standard_name="F57B10.7"
/notes="contains similarity to Pfam domain PF01204
(rehalase); coded for by the following C. elegans cDNAs:
yk86a7.5, CESE59F, yk195g8.3, yk195g8.5, yk375a12.5,
U51925"
/codon_start=1
/product="Hypothetical protein F57B10.7"
Query Match 18.7%; Score 321.8; DB 3; Length 36296;
Best Local Similarity 65.0%; Pred. No. 8.6e-60;
Matches 530; Conservative 0; Mismatches 242; Indels 43; Gaps 2;
QY 667 AAGCCACCGTTCATAGGCTGCGATGTTGTTAGAGCGGATATGCTCAATCTGAGACT 726
DB 22974 AAAAAATATCATTTTCCAGGTGTTGAAGTTGTTCCGAAGCGTTACGCTGCTGATGAACA 23033
QY 727 GACGAATTTCTGAACCAATGTTTTCGGAGCAT--GGGCGAGTAAAGATGACGAT 783
DB 23034 GACGAATTTGAAAGCAATCATCTTCAGGAGGAAAGAGCGGTGTTCAAAATGACGAT 23093
QY 784 ACTCTTTATTTCTCAATATGCTGCTGATCGTCAATTTGTTGAATGTTTGGGT 843
DB 23094 ACAATCATCTCTTCGACTACCGTCTGATCGTATCGGTGAGATTCTCGACCAATGGA 23153
QY 844 CTCGAACGTTATAAGATCTTAATAGTTTCGTTCCACCCAAAAAATATTCAGATAGT 903
DB 23154 ATGGATCGTTACAAGGATTCGATTCGAAGTTAGCTCATCCATCAATCTTCAAGTATAT 23213
QY 904 GGGATCAACCAATFACAAAGATTTCCATTTCCATGCTTATCCACCTGTGACTCAT 963
DB 23214 GGATGACTCATACAAAGCGGATTCCTCATTAATTCATCAATGCTGTTCCTCCGCGCAGCATGAAC 23273

```

```

QY 964 ACTAATGTGCTTCTGAATGGCTTCTCTCAAGGAGTTACTCAATTTCTACTGTGGGAA 1023
DB 23274 AAAAAATGATTTGCTGAGTGGCTCGCGGACCAAAAGTTTCGCAATTTGATTTGGGAA 23333
QY 1024 ACTGAGAAGTATCTCTCATGTTACCTTCTCTTTAAATTTGGTGGTGGAGGTTCAATTCCAA 1083
DB 23334 ACCGAAAAATACGCTCAGTTACATTTTCTTCAATGGAGGACTTGAAAAACAATTTGAG 23393
QY 1084 GATGAGAGCGTTGATGTTCCGTCACCAAAAG----- 1117
DB 23394 GGAGAAGAAGGTGTTTGTAGTCCCACTCCAAAGGTAATATTTGTGATGATATAACTTCGAA 23453
QY 1118 -----AAGTTGCTACATATGATTTAAAAACCAAGAAATGAATGCTGCTCGAGT 1163
DB 23454 AATTGAATATTTCCAGGTGCGAACTTACGATCTTCAACCAAGAAATGCTCGGCGCGCGT 23513
QY 1164 TCCGAAAAAATGTTGTCAGCAAAATGAGTCAGCAGCAGGATCCTTTGGTTATGTGCAATTT 1223
DB 23514 TGCTGACAAAATGATTTGAACAACCTCGAGGCTGGAACATCATCTTATTTGTCGAATTT 23573
QY 1224 TGGCCTCTCTGACATGTTGGACATACTGTAATTTGAACCTGCGCTCAAGCATGTCA 1283
DB 23574 TGCTCCACCAAGATATGTTGGGATACGGGAGTCTATGAAGTGTCTGTTCAAGGCTGTGA 23633
QY 1284 AGCTACTGACGAGGCAATTTGAAAGATATTTGAAGCATGCGCAAACTTATAATTTACCTTCT 1343
DB 23634 AGCTACTGATATCGCAATCGGAAGATCTATGAAGCACTCAAAAGCATGCTACTACT 23693
QY 1344 TATGTTACTTCCGATCATGGAATGCTGAGAGATGATGTTCCGATGGTAGTGAACA 1403
DB 23694 TATGTTACTTCTGATCAGGAAATGCTGAAAGATGAAGGCTCCAGATGGTGAACAACA 23753
QY 1404 TACTGCACATCTGCAATTTGTCCTCCATTTACTT 1438
DB 23754 CACTGCTCACATGTTACCGTGTTCCTCACTACTT 23788

RESULT 4
AL596172/c
LOCUS Listeria innocua CLIP11262 complete genome, segment 10/12.
DEFINITION AL596172 AL592022
ACCESSION AL596172.1 GI:16414852
VERSION
KEYWORDS
SOURCE Listeria innocua
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1
AUTHORS
Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A.,
Baquero,F., Berche,P., Bloeker,H., Brandt,P., Chakraborty,T.,
Charbit,A., Chetouani,F., Couve,E., de Daruvar,A., Dehoux,P.,
Domann,E., Dominguez-Bernal,G., Duchaud,E., Duranc,L.,
Dussurget,O., Entian,K.D., Fsihi,H., Portillo,F.G., Garrido,P.,
Gautier,L., Goebel,W., Gomez-Lopez,N., Hain,T., Hauf,J.,
Jackson,D., Jones,L.M., Kaerst,U., Krefit,J., Kuhn,M., Kunst,F.,
Kurapkat,G., Madueno,E., Maitournam,A., Vicente,J.M., Ng,E.,
Nedjari,H., Nordisiek,G., Novella,S., de Pablos,B., Perez-Diaz,J.C.,
Purcell,R., Rammel,B., Rose,M., Schlueter,T., Simoes,N.,
Tierrez,A., Vazquez-Boland,J.A., Voss,H., Wehland,J. and Cossart,P.
Comparative genomics of Listeria species
Science 294 (5543), 849-852 (2001)
21537279
PUBMED 11679669
REFERENCE
2 (bases 1 to 248050)
AUTHORS
Glaser,P., Frangeul,L. and Rusniok,C.
Direct Submission
Submitted (09-JUL-2001)
Microorganisms Pathogenes, 25 rue du Docteur Roux, 75724 Paris
Cedex 15, FRANCE
E-mail: pglaser@pasteur.fr
Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.
Location/Qualifiers
1. 248050
source

```



```

CDS
/gene="lin2337"
/9467..11392
/gene="lin2337"
/notes="similar to NADH oxidase"
/codon_start=1
/transl_table=11
/protein_id="CAC97564.1"
/db_xref="GI:16414860"
/db_xref="SPTREMBL:O929E0"
/translation="MKFNSMFSIDIGPVRNFRFVSPMCNNYANTDGLTDTSLAY
YKRALGGFLITFEATVVDVRANKGANKACLYSDHQIASFKRVIDACHDAGAKISVQ
LQHAGPEGNSGVYPLKASATAASAGRNTPEAISREELYELIYGLYEAALRAKAG
ADAEVCHVSGYLSFSGRTNRVDEFGCFENRMRLPRLIESIRKRVGHSIAIL
CRINSDVDGSLVDSQATVAAYLEDQGLDGHVSRVHIREDYVWAPVTLHAGSS
DLVTOIKRAVSIPIVIGRTEPHYAELMVRGRADLVAFROSADPPTPNKAAGK
LDELLPCIALQOCVANMYAGKPICTCLVNPPLGSEAYLPKTPSKVVYVIGGVGGL
YAGWAGSRHDIVYIETASDIIGQMRLAAYPPKGLDNLNMYRSYIKKEQDFVEIKT
NPTVPELQIDPADVITATGATPLVLPPIGDESLGIHAVDLDDKESCKRKLAV
GGWVGSETAAFLEAGHDVTVVELDEVGADVISHRKFLMRDFEYIKISITNAKV
TSFLPGDVTVSLADEKEYRIDGFSVVLAMGSRAYNPLEAIIKIVPVTVIGDAVRA
Query Match
Best Local Similarity 18.38; Score 314.2; DB 1; Length 248050;
Matches 767; Conservative 0; Mismatches 648; Indels 24; Gaps 4;
14 AGTTTGATGGAGCAATAATCAAAATGTTCAACAAAAGTCTGCTGTGTAGTTATTGATG 73
Db 214453 ATTATTCGAGCGAGTAATAATAATGAGTAATACCTGTAGCAATATTATCTCGATG 214394
74 GATGGGGCCCTTCGATGAACACACGGGAATCAATTCCTAAAGCTAAACGCCCTATTA 133
Db 214393 GTTTTGTAAACGTGCAGAACACAGTAGTAATGCTGTAGCTCAAGCAACAAACCAAA 214334
134 TGACAAACCTTCCTCGGAATTTGGCA --- AAAATTGGAGACACACGGCTTCATGTTG 190
Db 214333 TCAGCCGTTACTGGGCTAATTTCTCCACGGGAACCTTAAAGCTGTGGCTTTGATGTTG 214274
191 GATTGCCAGAGCTTAATGGGAATTCGAAGCTGGACATTTGAATATAGGAGCTGGAA 250
Db 214273 GTCTTCAGAGGCCAAATGGGTAACTCGAAGCTGGCCATAACAACATCGGAGCTGGAC 214214
251 GAGTATTATCAAGATATTGTTGCAATTAATTTGGCTGTTTCAACGAACGAGTTGTTA 310
Db 214213 GTATTGCTACCAAGCTTAACCTGATTGATGAAGCAATTGAAGAGCGCAATTCGAAG 214154
311 CAAATCCTCAGATTGTTGCATCAGCTGAGCGTCAAGAGGGGAGTGGTTCGATTGCATT 370
Db 214153 AGACAAAGCCCTAAACAATGCTTTCACTCATACAAAGAAACAATTCGACCTACATC 214094
371 TATTAGGACTGGTTAGCGATGGTGGTCCACTCTCATATTGATCATCTTTTTCGCTTGA 430
Db 214093 TTTTCGGTTTACTATCAGAGCGCGCTGCATAGTCACATTAATACCTCGTTGCATT 214034
431 TACGTGCATTTAAACAATTACAAGTCCGAAAGTTTTCATCTACTTTTTCGATGGTGC 490
Db 214033 TAGAAACTCGCAAGATAAAGGCGTGAAAACGCTTTATATCCATGCAATCTCTTGGAC 213974
491 GAGATACTTCGCCAACAAGTGGAGCTGTTATCTCAACAACCTTCCTCAATTTATTGCTT 550
Db 213973 GTGACGTGGCACCAACATCTCTACTAGATATTTAGAACACATAGAAAAGCTATTAGTG 213914
551 CGGAAAGTACGGAGAAATGGCTTACTATTACTGACGCTTATTATGCAATCGATAGGAC 610
Db 213913 ACTTAAACTATGTGCAATTTGCTTCTGACGCTTCTACGCGATGGATCGTGATA 213854
611 AAGATGGGAGCGTATTAGAGCGCTTATGAGGCAATTTGTTGAGGATTTGACAAAAAG 670
Db 213853 AACGCTGGGAACGTGTGAAAAGACATCAAAAGCAATCGTAAGCGCTGAAGGTGAAAAA- 213795
671 CCACCGCTGTATAGGCTGTGCGATGTTGTTAGAGCGGATATGCTCAATCTGAGACGTGCG 730
Db 213794 -----TTTGAGATCCAACTGCAAGCTTCTTACGCTTAATGACAAAAATGATG 213740
```

```

QY 731 AATTTCTGAAACCAATTGTTTTTTCGGACGATGGC-----GAGTAAAGATG 778
Db 213739 AATTCGTGTTCTCTGCTATCATTTACTTAAAGATGGCAACCTGTTGCAAGAGTTAAAGACA 213680
QY 779 ACGATACCTCTATTCTTCTCAATATCGTGTGCTGATCGTATGCGCTCAAAATTTGTAATGTT 838
Db 213679 ACGATGCAAGTGAATTTCTTCAATTTCCGCTCGACGCTGGATCAACTTCTTCAATGCAAT 213620
QY 839 TGGGCTCGAAGCCTTATAAGATCTTAATAGTTTCGTTCTCTCCTCACCTCAAAATATTTCAGA 898
Db 213619 TCACTGATAAAGAAATGGGATCATTTTCGACGCTGGGCTAACCACTTAAACATCAAAAT 213560
QY 899 TTAGTGGGATGACCAATCAATAAAGAGTTTCCATTTCCATCTGTTATTTCCCACTGTGA 958
Db 213559 TCGTTAATGACTCTTTACAAACCAAGCATTTGATGCTGAAGTTGCTTTGAGCCCAATTG 213500
QY 959 CTCATACCTAATGCTGTGTAATGCTGCTTCTCAAGAGAGTTACTCAATTTCACTGTG 1018
Db 213499 AATGAAAAATGTTATCGCGGAAGTACTTTCTAATGAAGGTTTGTCCGAACCTCGTATCG 213440
QY 1019 CGAAACTGAGAAGTATCTCTCATGTTTACCTTCTCTTCTTAAATGCTGCGAGAGTTCAAT 1078
Db 213439 CTGAAACAGAAAAATATCCACAGTAAACATTTCTTTATGAATGCGGACGAAATGAAGAAT 213380
QY 1079 TCCAAGATGAAGAGCTGTTGATGTTCCGTCACCAAAAGAGTTGCTACATATGATTTAA 1138
Db 213379 TCCTCGTGAAACCGGATCTTAATCAATTCGCAA---AAGTAGAAACATACGATTTAA 213323
QY 1139 AACCAAGAAATGAATGCTGCTGGAGTTGCCGAAAAAATGTCGAGCAAAATTTAGTCAGGCA 1198
Db 213322 AACCTGAAATGAGTGCATATGAAGTAACAGATCGCTGTTGTAAGACATTTAAAAACGACA 213263
QY 1199 GGCATCTTGGTTATGTCGAATTTTGGCCCTCTGACATGTTGACATCTGCTAAAT 1258
Db 213262 AACAGATCGGATCATCTTAACTTCGTAATCCAGATGTTGTCATCTGAGCTATGTC 213203
QY 1259 TTGAACCTGCCGTCAAGCATGTCGAAGCTACTGACGAGGCAATTTGGAAGATATTTGAAG 1318
Db 213202 TTGAGCCAACTATTAAAGCAATCGAAGCAGTAGATGAAATCTTGTCTGTAGTAGATC 213143
QY 1319 CATGCCAAACTTATAATTAGTCTTCTTATGTTTCTTCCGATCATCAATTTGCTGAGAGA 1378
Db 213142 TTATTTTGAAGAAAGGTGCTCAGCTATTATCTTCTGCTGACCATGTTACTCTGAAACAA 213083
QY 1379 TGATTCTCCCGATGTTAGTGAACATCTGACATACCTGCAATTTGGTCCCAATTTACT 1437
Db 213082 TGTCTACTCCAGAGGAAACCCGACACTGCGCACACTACCGTTCAGTTCCAGTAAT 213024
```

```

RESULT 5
AX413018/c 349980 bp DNA linear PAT 02-SEP-2002
LOCUS Sequence 9 from Patent WO0228891.
DEFINITION AX413018
ACCESSION AX413018
VERSION AX413018.1 GI:21445476
KEYWORDS
SOURCE Listeria innocua
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1
AUTHORS Kunst, F. and Glaser, P.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 9 11-APR-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
Location/Qualifiers
source
1. 349980
/organism="Listeria innocua"
/mol_type="genomic DNA"
/db_xref="taxon:1642"
/note="old seq 9 (684707) splitted in 3-new seq 9 = old
seq 9 from 1 to 349980-seq 4027 = old seq 9 from 300000 to
649980-seq 4028 = old seq 9 from 600000 to 684707"
```


BASE COUNT	104581 a	69113 c	62414 g	113872 t	
ORIGIN					
Query Match	18.3%	Score 314.2;	DB 6;	Length 349980;	
Best Local Similarity	53.3%	Pred. No. 2.7e-58;			
Matches 767;	Conservative 0;	Mismatches 648;	Indels 24;	Gaps 4;	
QY	14	AGTTTGAGTGGACAAATATCAAAATGTTCAACAAAAGTCTGTCTTGTAGTTATTGATG	73		
DB	138870	ATTATTGGAGCAGTAATAATAATGATGAATACACCTGTAGCAATATTATTATCTCCGATG	138811		
QY	74	GATGGGCTTTCCGATGACACACACCGGAATGCAATTTGCTTAAAGCTTAAACGCGCTATTA	133		
DB	138810	GTTTTGTAAACGTCGAGAAACAGTAGTAATGCTGTAGCTCAAGCAACAAACAAAT	138751		
QY	134	TGGCAAACTTTGCTTGGAAATGGCA--AAATTTGGAAGCACACGGCTTTCATGTTG	190		
DB	138750	TCGACCGTTACTGGCTTAATTTCTCAGGGGAACCTTAAGCTGTGGCTTGATGTTG	138691		
QY	191	GATTCGACAGAGGCTTAATGGGAAATTCGAAATTTGGACATTTGAATATAGGAGCTGGAA	250		
DB	138690	GTCTTCAGAGGCAAAATGGGTAACTCTGAATTTGGCCATACAAACATCGAGCTGGAC	138631		
QY	251	GAGTTATTATCAAGATATTGCTCGAATTAATTTGGCTGTTCACGAAACAGAGTTGTTA	310		
DB	138630	GTATTGCTACCAAGCTTAACTCGTATTGATAAGCAATTTGAAGAAGCGCAATTCGAAG	138571		
QY	311	CAAACTCTCAGATTTGTCATCAGCTGAGCGTCAAAAGAGGGAGTGGTGATTGATT	370		
DB	138570	AGACAAAGCCCTAAACATGCTTTCATCTACATACAAAGAAACAAATTCGACCTACATC	138511		
QY	371	TATTAGGACTGGTTAGCGATGGTGGTGTCCACTCTCATATTGATCATCTTTTTCGGTTGA	430		
DB	138510	TTTTCGGTTTACTATCAGACGGCGGTGCATAGTACATTAATCACTCTGTCACCTTT	138451		
QY	431	TACGTGATTTAAACATTTACAAGTGCACAAAGTTTTCATTCACCTTTTTCGTCAGGTC	490		
DB	138450	TAGAACTGCGAAGATAAAGGCGTGAACACGTTTATATCCATGCAATTCCTTCATGGAC	138391		
QY	491	GAGTACTTCCGCAACAAAGTGAGCTGGTATTCTTGAACAACTCTTCAATTTTATGCTT	550		
DB	138390	GTGACGTGGCACCAATCTCCTACATAGATATTTAGAAACACTAGAAAACCTATTAGTG	138331		
QY	551	CGGAAAAGTACGGAGAATTGGCTACTATTACTGACGCTTATATGCAATGATGAGGACA	610		
DB	138330	ACTTAAACTATGTGCAATTCCTACTGTTCTGCACTGTTTCTACGCGATGATGCTGATA	138271		
QY	611	AAAGATGGGACGTTTAAAGTGGCTTATGAGGCAATTTGTTGAGGTATTGGACAAAAG	670		
DB	138270	AACGCTGGAAACGTTTGAAGAAACATACAAAGCAATCGTAAGCGCTGAAGGTGAAAAA	138212		
QY	671	CCACCGTTGATAAGGCTGTGATGTTGTTAGAGGCGATATGCTCAATCTGAGACTGACG	730		
DB	138211	-----TTTGAATATCCAACTCACTGTTGCAAGCTTCTAGCCTAAATGACAAAATGATG	138157		
QY	731	AATTTCTGAAACCAATTTGTTTTTCGGACGATGGGC-----GAGTAAAGATG	778		
DB	138156	AATTCGTTGTTCTGCTATCATATTACTAAAGATGCAACACCTGTTGCAACATTTAAAGACA	138097		
QY	779	ACGATCTCTTATTTTCTTCAATTTATGCTGCTGATGCTGATGCTGCTCAATTTGTTGAATGTT	838		
DB	138096	ACGATGACGATGTTTCTTCAATTTCCGCTCCGCTACCGCTGGGATTCACCTTTCTATGAT	138037		
QY	839	TGGTCTCGAAGCTTATAAGATCTTAATAGTTCGTTCCCTCACCTTAAATAATTCAGA	898		
DB	138036	TCACGTATAAGAAATGGATCATTTTCGACCGTGGGCTTAACCCCTTAAACATCAAT	137977		
QY	899	TTAGTGGGATGACCAATACAAATAAGAGTTTCCATTTCCATCTGTTATTCACCACTGTGA	958		
DB	137976	TCGTTTACAAATGACTCTTTACAAACCAAGCAATGATGCTGAAGTTGCTTTTTCAGCCAA	137917		
QY	959	CTCATACTAATGTGCTGCTGAATGGCTTCTCAAGGAGTTTACTCAATTTTCACTGTG	1018		
DB	137917	-----			
DB	137916	AAATGAAAAATGTTTATCGGCAAGTACTTTTCTAATGAAGTTTTCGCAACTCGCTATCG	137857		
QY	1019	CGGAACTGAGAACTATCCCTCATGTTTACCTTCTTCTTAAATGTTGGTGGAGAACTTCAAT	1078		
DB	137856	CTGAACAGAAAAATATCCACACGTAACNTTCTTTTATGAATGGCGGAGAAATGAAGAT	137797		
QY	1079	TCCAAAGTGAAGACGCTTGTATGTTCCGTCACCAAAAAGAGTTGCTACATATGATTTAA	1138		
DB	137796	TCCCTGTGAAACCGGATCTTAATCAATTCGCCAA--AAGTAGAAACATACGATTTAA	137740		
QY	1139	AACAGAAATCAATGCTGCTGGAGTTGCCGNAAAAATGGTCGACCAATTTGAGTCAGCA	1198		
DB	137739	AACCTGAATGAGTGCATATGAAGTAACAGATGCGCTTGTGTAAGACATTTAAAAACGACA	137680		
QY	1199	GGCATCTTGGTTATGTCGAATTTTGGCCTCTCGATCATGTTGGACATCTGCTAAAT	1258		
DB	137679	AACAGATGCGATCATCTTAACCTTCGTAATCAGATATGTTGGACATTCAGGATGTC	137620		
QY	1259	TTGAACCTGCGCTCAAAAGCATGTCAGCTACTGACGAGGCAATTTGGAAGATATTTGAAG	1318		
DB	137619	TTGAGCCAACCTATTAAAGCAATCGAAGCAGTAGATGAAATCTTTGGTCGTAGTAGATC	137560		
QY	1319	CATCCCAACTTATAATTACGTTCTTATGTTTACTTCCGATCATGGAATCTGAGAGA	1378		
DB	137559	TTATTTAGAAAAAGTGGCTCAGCTATTATCTTTGCTGACCACTACCGTTCCAGTTCCAGTAAT	137500		
QY	1379	TGATTGCTCCGATGTTAGTGAACATCTGCACATACCTGCAATTTGTCCTCCATTTACT	1437		
DB	137499	TGCTACTCCAGAGGAAACCGCACACTGGCACACTACCGTTCCAGTTCCAGTAAT	137441		
RESULT 6					
AX417048/c					
LOCUS	AX417048	349980 bp	DNA	linear	PAT 02-SEP-2002
DEFINITION	Sequence 4039 from Patent WO0228891.				
ACCESSION	AX417048				
VERSION	AX417048.1	GI:21449658			
KEYWORDS					
SOURCE	Listeria innocua				
ORGANISM	Listeria innocua				
REFERENCE	1				
AUTHORS	Kunst,F. and Glaser,P.				
TITLE	Listeria innocua, genome and applications				
JOURNAL	Patent: WO 0228891-A 4039 11-APR-2002;				
	INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE				
	SCIENTIFIQUE (CNRS) (FR)				
FEATURES	Location/Qualifiers				
source	1. 349980				
	/organism="Listeria innocua"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:1642"				
	/note="seq 2058, original length: 3.011.208 replaced				
	by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to				
	0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034;				
	0.900.001 to 1.249.980-seq 4035: 1.200.001 to				
	1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037;				
	1.800.001 to 2.149.980-seq 4038: 2.100.001 to				
	2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040;				
	2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"				
BASE COUNT	101702 a	72631 c	57863 g	117784 t	
ORIGIN					
Query Match	18.3%	Score 314.2;	DB 6;	Length 349980;	
Best Local Similarity	53.3%	Pred. No. 2.7e-58;			
Matches 767;	Conservative 0;	Mismatches 648;	Indels 24;	Gaps 4;	
QY	14	AGTTTGAGTGGACAAATATCAAAATGTTCAACAAAAGTCTGTCTTGTAGTTATTGATG	73		
DB	176453	ATTATTGGAGCAGTAATAATAATGAGTAATCACCTGTAGCAATATTATTCTCCGATG	176394		
QY	74	GATGGGCTTTCCGATGAAACACCGGGAATGCAATTTGCTTAAAGCTTAAACGCTATTA	133		

Db	176393	GTGTTGGTAAACGCTGCAGAACAGTAGTAATGCTGTAGCTCAAGCAAAACCAAAAT	176334
QY	134	TGGACAAACCTTCTCTCGAAATGGCA---AAATTTGGAAGCACACGGCTCTCATGTTG	190
Db	176333	TCGACCGTTACTGGGTAAATTTTCCACGGGAACCTTAAAGCTGCTGGCTTGTATGTTG	176274
QY	191	GATTGGCAGAGCTTAATGGGAAATTCGAAGTTGGACATTTGAATATAGGAGCTGGAA	250
Db	176273	GTCTCCAGAGGCCAAATGGTAACCTGGAAGTTGGCCATACAAACATCGGAGCTGGAC	176214
QY	251	GAGTATATCAAGATATTTGTCGAATTAATTTGGCTGTTCAACGAAGAGATTGTTA	310
Db	176213	GTATTGTTACCAAGCTTAACCTGATTTGATTAAGCAATTAAGAGGCGCAATTCGAAG	176154
QY	311	CAATCTCAGATGTTGTCATCAGCTGAGCGTCAAGAGAGGGAGTGGTCGATTCGATT	370
Db	176153	AGACCAAGCCCTAAACAATGCTTTCATCTATACAAAGAAACAATTCGGACCTACATC	176094
QY	371	TATTAGACTGGTTAGCGATGGTGTGCCACTCTCATATTTGATCATCTTTTTCGCTTGA	430
Db	176093	TTTTCGGTTTACTATCATCAGCGCGCGCTGCATAGTCACATTAATCACCTCGTTGCAC	176034
QY	431	TACGTGCATTAACGAATTACAAGTCCCAAGGTTTTCATTCACCTTTTTCGTGATGGTC	490
Db	176033	TAGAACTCGCAAGATAAAGCGCTGAAAACGTTTATATCCATGCAATTCCTTGATGGAC	175974
QY	491	GAGATACTTCGCCCAACAAGTGGAGCTGTTATCTTGAACAACCTCTTCAATTTATGCTT	550
Db	175973	GTACGTGGCACCCACATCTCCTACTAGATATTTTAGAACACTAGAAAAGCTATTAGTG	175914
QY	551	CGAAAGATACGGAGAAATGGCTPACTTACTGACGCTTATTATGCAATGGATAGGACA	610
Db	175913	ACTTAAACTATGTGCAATTTGCTACTGTTTCTGACGTTTCTACGCGATGGATCGTGATA	175854
QY	611	AAAGATGGGAGCTATTAAAGATGGCTATGAGCAATTTGCGAGTATTTGACAAAAG	670
Db	175853	AACGCTGGGACGTTGTGAAAAGCATACAAAGCATCGTAAGCGCTGAAGGTGAAAAA	175795
QY	671	CCACCGTTGATAGGCTGTCGATGTTGTTAGAGAGCGATATGCTCAATCTGACACTGAGC	730
Db	175794	-----TTTGAAGATCCAATCGAACTTGTCAAGCTTCTTACGCTATGACAAAATGATG	175740
QY	731	AATTTCTGAACCAATGTTTTTTCGGACGATGGGC-----GAGTAAAGATG	778
Db	175739	AATTCGTTCTCTGCTATCATTAATAAGATGGCAAAACCTGTTGCAACAGTTAAAGACA	175680
QY	779	ACGATACTCTTATTTCTCAATATCGTGCATCGTATGCTCAAAATTTGTAATGTT	838
Db	175679	ACGATCGAGTATTTCTCAATTTCCGTCGACGCGGATTCACCTTTCTAATGCAAT	175620
QY	839	TGGGCTCGAACGCTTATAAGATCTTAAATAGTTCGGTTCTCTCACCCCTAAAAATATTCA	898
Db	175619	TCACTGATAAAGATGGATCAATTCGACGCTGGGCTAACCCACCTAAAAACATCAAT	175560
QY	899	TTAGTCGGATGACCCAAATACAATAAGATTTCCATTTCCATGTTATTTCCACCGTGA	958
Db	175559	TCGTTACAATGACTCTTTACAACCCCAAGCATTTGATGCTCAAGTTGCTTTGAGCCAA	175500
QY	959	CTCATACTAATGCTGCTGAATGCTTCTCTCAAGSAGTTACTCAATTTCTACTGTG	1018
Db	175499	AAATGAAAATGTTATCGCGAAGTACTTTCTAATGAAGGTTTGTGCAACTGCGTATCG	175440
QY	1019	CGAAACTGAGAAGTATCTCTATGTTTACCTTCTTTAATGGTGGTGCAGAAATTTCAAT	1078
Db	175439	CTGAACAGAAAATATCCACAGTAACATTTCTTTATGAATGGCGACCAATGAAGAT	175380
QY	1079	TCCAAGATCAAGCGCTGTATGTTTCCGTACCAAAAGATTTGCTACATATGATTAA	1138
Db	175379	TCCCTGGTGAACCGGATTTCTAATCAATTCGCCAA---AAGTAGAAACATACGATTTAA	175323
QY	1139	AACCCAGAAATGATGCTGCTGGAGTTGCGGAAAAAATGTCGAGCAAAATTTAGTCAGGCA	1198
Db	175322	AACCTGAATGATGCTATGAAGTAAACAGATGCGGCTTGTGAAGCATTAACAAAGACA	175263
QY	1199	GCATCCTTTGTTATGTGCAATTTTTCGCCCTCCTGCATGTTGGACATCTGGTAAAT	1258
Db	175262	ACACGATCGGATCATCTTAAACTTCGCTAAATCCAGATATGTTGGACATTCAGGTATGC	175203
QY	1259	TTGAACCTCCGCTCAAAGCATGTCAGCTACTGACGAGGCAATTTGAAAGATATTTGAAG	1318
Db	175202	TTGAGCCAACTATTAAAGCAATCGAAGCAGTAGATGAAAATCTTGGCTGTAGTAGATC	175143
QY	1319	CATGCCAACTTATTAATTAAGTTCATTATGTTTACTTCCGATCATGAAATGCTGAGAAGA	1378
Db	175142	TTATTTTGAAGAAAGGTGGCTCAGCTATATCTTGTCTGACCATGTTAACTCTGAAACAA	175083
QY	1379	TCATTCCTCCGATGTTAGTGAACATGTCACATACCTGCAATTTGGTCCCATTTACT	1437
Db	175082	TGCTACTCCAGAAAGGAAAACCGCACACTGGCACACTACCGTTCAGTTCAGTAAT	175024
RESULT 7			
AL591983/c			
LOCUS	AL591983	324050 bp	DNA linear BCT 06-JUN-2002
DEFINITION	Listeria monocytogenes strain EGD, complete genome; segment 11/12.		
ACCESSION	AL591983	AL591824	
VERSION	AL591983.1	GI:16411809	
KEYWORDS	Listeria monocytogenes		
SOURCE	Listeria monocytogenes		
ORGANISM	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.		
REFERENCE	1		
AUTHORS	Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K. D., Eshti, H., Portillo, F. G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L. M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkak, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E., Nedjari, H., Nordisiek, G., Novella, S., de Pablo, B., Perez-Diaz, J. C., Purcelli, R., Remmel, B., Rose, M., Schluter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J. A., Voss, H., Wehland, J. and Cossart, P. Comparative genomics of Listeria species		
TITLE	Science 294 (5543), 849-852 (2001)		
JOURNAL	21537279		
MEDLINE	11679669		
PUBMED	2		
REFERENCE	(bases 1 to 324050)		
AUTHORS	Glaser, P., Frangeul, L. and Rusniok, C.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE		
COMMENT	E-mail: pglaser@pasteur.fr Phone: +33 1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.		
FEATURES	Location/Qualifiers		
source	1..324050		
	/organism="Listeria monocytogenes"		
	/mol_type="genomic DNA"		
	/strain="EGD-e"		
	/db_xref="taxon:1639"		
gene	complement(290..492)		
CDS	/gene="lmo321"		
	complement(290..478)		
	/gene="lmo321"		
	/codon_start=1		
	/transl_table=11		
	protein_id="CAD00399.1"		
	/db_xref="GI:16411810"		
	/db_xref="SPTREMBL:Q928W7"		
	/translation="MAERIFRQTFGNSEIFIDDRTHIANPAFRQKIALIETGCEKMTDYTEELKLAGYEEVTR"		
RBS	complement(487..492)		
gene	/gene="lmo321"		
	complement(586..801)		
	/gene="lmo322"		

```
CDS      complement(586..801)
/gene="lmo2322"
/codon_start=1
/transl_table=11
/product="gp44 [Bacteriophage A118]"
/protein_id="CAD00400.1"
/db_xref="GI:16411811"
/db_xref="SPTREMBL:O8Y4V8"
/translacion="MSEVLVSASVEGYESKSFNFTEINNIVKERFKIDVERKKRAE
TFNKKEVTVKELVDGHLREIIPRRL"
/gene="lmo2323"
complement(join(798..1331,1340..1345))
/gene="lmo2323"
complement(798..1331)
/gene="lmo2323"
/codon_start=1
/transl_table=11
/product="gp43 [Bacteriophage A118]"
/protein_id="CAD00401.1"
/db_xref="GI:16411812"
/db_xref="SPTREMBL:O8Y4V7"
/translacion="MFQKSTSAAMQVLAETRTQKELADISVYTPALISNQIKGKRT
VSLEQABQLDSYNEPRSTYLFARHSNGMIPPLFDGLDNHSLTNRFEEVEEANN
TLKNGLTMTFNLKRGDMLQREAKQAISITDVSALTNASIAFAFNIDLOQVLN
KRDLYOKSGLVRSCKG"
complement(810..815)
/gene="lmo2322"
complement(810..815)
/gene="lmo2322"
complement(1340..1345)
/gene="lmo2323"
complement(1454..2230)
/gene="lmo2324"
complement(1454..2230)
/gene="lmo2324"
note="similar to anti-repressor [Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00402.1"
/db_xref="GI:16411813"
/db_xref="SPTREMBL:O8Y4V6"
/translacion="MSNLQVIANEMLPVLENEQGEKFDARMLEKLLVTRFNDWIY
RMIGNYENGLDFYSTLNGRSPSTNYFLTDATAKELAWQNNEMGRAIRKIFIEV
EKQAKLATYPAFYMIEDPVARAKKIEQEQKQVLEKLEQKPKVVFARAVQTS
ENTILVDLATILKGLDIGNRFLFWLRGSGVLLSKGAYNKPQKAMNGLIFEQK
THIHDRNGLMKRTYTPQVTGKGQVLLNKLLEHQNQV1"
complement(2212..2487)
/gene="lmo2325"
complement(2212..2487)
/gene="lmo2325"
/codon_start=1
/transl_table=11
/protein_id="CAD00403.1"
/db_xref="GI:16411814"
/db_xref="SPTREMBL:O8Y4V5"
/translacion="MIFTVNSFPQSGHVFPGLTVNTCEHSGQLSPVFLIISCFALF
RLCSLICLSNFFSNLILTPILPTSLKNTSTVKGRTERTKQIYK"
complement(2236..2241)
/gene="lmo2324"
complement(2236..2241)
/gene="lmo2324"
complement(join(2493..2774,2782..2787))
/gene="lmo2326"
complement(2493..2774)
/gene="lmo2326"
note="similar to protein gp41 [Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00404.1"
/db_xref="GI:16411815"
/db_xref="SPTREMBL:O8Y4V4"
/translacion="MNNIKAILKLEILENGNAIESGSFVKYSVKNILSLLEKDOE
LKTIEMKVLNGVEDSIENATLLEKRLUEAKSLVEDLASTINSLEIKVK"
complement(2498..2503)
/gene="lmo2325"
complement(2498..2503)
/gene="lmo2325"
complement(2782..2787)
/gene="lmo2326"
complement(2800..3097)
/gene="lmo2327"
complement(2800..3084)
/gene="lmo2327"
/codon_start=1
/transl_table=11
/protein_id="CAD00405.1"
/db_xref="GI:16411816"
/db_xref="SPTREMBL:O8Y4V3"
/translacion="MNRYSLNKRSEQKDCVFKTDIPPENLKSINIKMOGDRNCCYG
VLEINGQLRKGITAVKLDLKGSLPVVQVEYHPFTTISEMQRLWSGKY"
complement(3092..3097)
/gene="lmo2327"
complement(3287..3540)
/gene="lmo2328"
complement(3287..3529)
/gene="lmo2328"
note="similar to transcription regulator"
/codon_start=1
/transl_table=11
/protein_id="CAD00406.1"
/db_xref="GI:16411817"
/db_xref="SPTREMBL:O8Y4V2"
/translacion="MRIWLKIREENGLTQODVAELADVERTTYASTEQGRNPSVYN
AMRIASVLDFEWTFIFDDKVKREKSHFYLNKEQEAR"
complement(3535..3540)
/gene="lmo2328"
3672..4161
/gene="lmo2329"
3672..3677
/gene="lmo2329"
3685..4161
/gene="lmo2329"
note="similar to a putative repressor protein
[Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00407.1"
/db_xref="GI:16411818"
/db_xref="SPTREMBL:O8Y4V1"
/translacion="MFGDILMKLNKLTQSDIAKILGVARTTYSYEQNRMPDAE
IQIRADIYDVSLDYLDHGRKNNTADTIAAHIDNSAEEDEIKELATIEKREKHEVNE
EENITDAASKGDKVEVNEFDENEDFKAVARIMSDAEAVKAKVSFIEFYEQQKNN"
4528..5243
/gene="lmo2330"
4528..4533
/gene="lmo2330"
4542..5243
/gene="lmo2330"
note="similar to protein gp33 [Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00408.1"
/db_xref="GI:16411819"
/db_xref="SPTREMBL:O8Y4V0"
/translacion="MSKYSYLLKKWFWMILLILLISLFGNGIWLFFATLATLTFAI
IKVTKNENRRKYLTIITATVLLTLTSLIRTVOLYNYTVENPENVTIDNOKKOTKED
AKTDEVEETEDQADASAAYVTQEGIESFESINRILISFSGILIKVVPFENEYDM
LIAYVTODLAKYQDETIIKQNADYLGNEIQQRALGTFLGGDSNHRPVELRYEDETMA
GSSAFDRTNMLTKG"
5251..5944
/gene="lmo2331"
5251..5256
/gene="lmo2331"
5264..5944
/gene="lmo2331"
note="weakly similar to gp32_Bacteriophage A118 protein"
/codon_start=1
```

/transl_table=11

```
Query Match      18.2%; Score 312.6; DB 1; Length 324050;
Best Local Similarity 53.2%; Pred. No. 6e-58;
Matches 766; Conservative 0; Mismatches 649; Indels 24; Gaps 4;

QY 14 AGTTTGAGATGGCAAAATATCAAAATGTTTCAACAAAAGTCCTCTTTGTAGTTATTGATG 73
DB 14205 ATTATTGGAGGCGAGTAAATAAATAAATGAGTAATACCTGTGAGCAATATTATCTCCTCGATG 134146

QY 74 GATGGGCGCTTTCCGATGACACACGCGAATGCAATGCTAAAGCTAAACCCCTATTAA 133
DB 134145 GTTTTGTAAACGTGACGAAAGTAGTAGTAACTCTAGCTCAAGCAACAAACCAAAAT 134086

QY 134 TGCACAAAACCTTTTCTCGAAATTTGGCA---AAAATTGGAAGCACACGGTCTTCATGTTG 190
DB 134085 TCACCGTTATTGGCTGATTTCCTCACGGGAACCTTAAAGCTGTGCGCTTGATGTTG 134026

QY 191 GATTGCCAGAGCTTTAATGGGAATTTCTGAAGTTGGACATTTGAATATAGGAGCTGGAA 250
DB 134025 GCCTTCAGAAAGTCAAAATGGGTAACTCTGAAGTTGGCCATACAAACATCGGAGCTGGAC 133966

QY 251 GAGTTATTATCAAGATATGTTGCAATTAATTTGGCTGTTCAACGAACGAGTTTGTTA 310
DB 133965 GTATTGCTACCAAAAGCTTAACCTGATTGACAAAGCTATTTGAAGAAGCGCAATCCAAAG 133906

QY 311 CAAATCCTCAGATTGTTGCATCAGCTGAGCGTGCAAAAGAGGGAGTGGTGCATTGCAATT 370
DB 133905 AGAACAAAGCCCTAAACAATGCTTTTCACTCATACAAAGAACAACACTCGGACTTACATC 133846

QY 371 TATTAGAGCTGGTTAGCGATGGGTGTCACCTCTCATATTTGATCATCTTTTTCGCTTGA 430
DB 133845 TTTTTCGGTTTACTATCAGACGGCGGTGCTAGTACATTAATACCTCGTTGCACCTTT 133786

QY 431 TAGCTGCATTTAAACAATTAACAGTCCCAAGGTTTTCATCTACCTTTTTCGTGATGGTC 490
DB 133785 TAGAACTGGGAAGATAAAGCGGTGAAAACGCTTTATATCCATGATTCCTTTGAGGAC 133726

QY 491 GAGATCTTCGCCAACAGTGGAGCTGGTTATCTTGAACAACCTTCCTCAATTTATTGCTT 550
DB 133725 GTACGCTGGCACCACAACTCTACTAGATAATTTTGAACAACACTACAAAAGCTATTAGTG 133666

QY 551 CGGAAAGTACGAGAAATTTGGCTACTATTACTGGAGCTTATTATGCAATGATAGGACGA 610
DB 133665 ACTTAAATCATGCGCGGATTTGCTACTGTTCTTGACGCTTCTACGCGATGGATCGTGATA 133606

QY 611 AAAGATGGGAGCTATTAAAGATGGCTTATGAGGCAATTTGTTGAGGTATTGGACAAAGAG 670
DB 133605 AACGCTGGGAACGTTGTGAAGAAGCATACAAAGCAATCTGACGCTGAAGGTGAAGAAA- 133547

QY 671 CCACCGTTGATAGGCTGTCGATGTTGTTAGAGAGCGATATGCTCAATCTGAGACTGACG 730
DB 133546 -----TTTGAAGATCCAATCGAACTTGTCAAAGCTTCTTATGCTATGACAAAATGATG 133492

QY 731 AATTTCTGAACCAATGTTTTTTCGGAGCATGGGC-----GAGTAAAGATG 778
DB 133491 AATTCGTTCTCTGCTATCATTTACTATAAGATGGCAACCTGTTGCAACAGTTAAAGACA 133432

QY 779 ACGATCTCTTATTTCTCAATATCGTGGCTGATGCTGATGCTCAAAATTTTGTGATGTT 838
DB 133431 ACGATGCACTTATTTCTCAATTTCCGCTCTGACCGCGATTAACCTTTCTTAATGCAAT 133372

QY 839 TGGGTCTCGAAGCTTATAAGATCTTAATAGTTTCGCTTCCCTCACCCCTAAAAATATTACA 898
DB 133371 TCACTGATAAAGATGGGATCATTTTCGACCGTGGGCGAGATCACCTAAAAACATCAAT 133312

QY 899 TTAGTGGGATGACCCATACATAAAGATTTCCATTTCCATCGTTATTTCCACCTGTA 958
DB 133311 TTGTTACAATGACTCTTTTACACCCCAAGTATTGATGCTGAAGTTGCTTTTGAGCCAAATG 133252

QY 959 CTCATCTAATGCTGCTGATGCTGCTTCTCAAGGAGTTACTCAATTTCACTGTTG 1018
DB 133251 AATGCAAAATGTTATCGCGGAAGTACTTTCTAATGAAGGCTTATCGCAACTGCGTATCG 133192
```

```
QY 1019 CGAAACTGAGAAGTATCCTCATGTTACCTTCTTCTTAATGTTGGTGCAGAAATTCAT 1078
DB 133191 CTGAACACAGAAATATCCACAGTAACATTTCTTTATGATGTTGGCAGCAATGAAGAT 133132

QY 1079 TCAAGATGAAGAGCGCTTGTATGTTCCGTCACCAAAAGAGTTGTACATATGATTTAA 1138
DB 133131 TCCTGTTGAAAACCGGATTTCTAATCAATTCGCAA---AAGTAGAAAACATGCAATTTGC 133075

QY 1139 ACCAGAAATGATGCTGCTGGAGTTGCCGAAAAAATGTCGAGCAAAATGAGTCAGGCA 1198
DB 133074 AACCTGAAATGAGCGCATATGAAGTAACAGATGCACTTGTGTAAGACATTTAAAAAGCA 133015

QY 1199 GGCATCCTTTGGTTATGTCGAATTTTGGCCTCTGCACATGTTGGACATCTGCTAAAT 1258
DB 133014 ACACGATCGGATCATCTTAACTTCGCAACCCAGATATGTTGGGCACTCAGGTATGC 132955

QY 1259 TTGAACCTGCCGTCAAAGCATGTCAGCTACTGACGAGGCAATTTGAAAGATATTTGAAG 1318
DB 132954 TTGAGCCAACGATTAAAGCAATCGAAGCAGTAGATCAAAATCTTGGCTGTAGTAGACC 132895

QY 1319 CATGCCAACTATATTAATAGTCTTATGTTACTTCCGATCATGCAATGCTGAGNAGA 1378
DB 132894 TTATTTTGAAGAAAGTGGTTGAGCTATTATCTTGTGCTGACCATGTAATCTGGAACAA 132835

QY 1379 TGATTCCTCCGATGCTAGTGAACATPACTGCACATACCTGCAATTTGGTCCCATTTACT 1437
DB 132834 TGCTACTCCAGAGGAAAACCCGACACTGCGCACACTCCGTTCCAGTCCAGTAATT 132776
```

RESULT 8

```
AX641672/c
LOCUS      AX641672      349980 bp      DNA      linear      PAT 21-FEB-2003
DEFINITION Sequence 2862 from Patent WO0101118.
ACCESSION  AX641672
VERSION     AX641672.1  GI:28474433
KEYWORDS
SOURCE      Listeria monocytogenes
            Listeria monocytogenes
ORGANISM    Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE   1
AUTHORS     Buchrieser,C., Frangeul,L., Couve,E., Rusniok,C., Psihi,H.,
            Dehoux,P., Dussurget,O., Chetoui,F., Nedjari,H., Glaser,P.,
            Kunat,F., Cossart,P., Daniels,J., Goebel,W., Krefte,J., Kuhn,M.,
            Ng,E., Vasquez-Boland,J., Dominguez-Bernal,G., Garrido-Garcia,P.,
            Tierrez-Martinez,A., Amend,A., Chakraborty,T., Domann,E., Hain,T.,
            Berche,P., Charbit,A., Durant,L., Perez-Diaz,J.C., Baquero,F.,
            Garcia del Portillo,F., Gomez-Lopez,N., Madsenio,E., de Pablo,B.,
            Wehland,J., Kaerst,U., Entian,K.D., Hauf,J., Rose,M. and Voss,H.
            Listeria monocytogenes genome, polypeptides and uses
            Patent: WO 0101118-A 2862 11-APR-2001;
            INSTITUT PASTEUR (FR)
```

FEATURES

source

```
1. 349980
/organism="Listeria monocytogenes"
/mol_type="genomic DNA"
/db_xref="taxon:1639"
/note="Seq 1 for big: original length: 2944528-split
into 10 <23>seq 0001 : From 0.000.001 to 0.349.980
length: 349.980 <23>seq 2855 : From 0.300.001 to
0.649.980 length: 349.980 <23>seq 2856 : From 0.600.001
to 0.949.980 length: 349.980 <23>seq 2857 : From
0.900.001 to 1.249.980 length: 349.980 <23>seq 2858 :
From 1.200.001 to 1.549.980 length: 349.980 <23>seq 2859
: From 1.500.001 to 1.849.980 length: 349.980 <23>seq
2860 : From 1.800.001 to 2.149.980 length: 349.980
<23>seq 2861 : From 2.100.001 to 2.449.980 length:
349.980 <23>seq 2862 : From 2.400.001 to 2.749.980
length: 349.980 <23>seq 2863 : From 2.700.001 to
2.944.528 length: 244.528"
BASE COUNT 101075 a 74760 c 61320 g 112825 t
ORIGIN
```

Query Match	18.2%; Score 312.6; DB 6; Length 349980;
Best Local Similarity	53.2%; Pred. No. 6e-58;
Matches	766; Conservative 0; Mismatches 649; Indels 24; Gaps 4;
QY	14 AGTTTGAGATGGACAATATCAAAATGTTTCAACAAAAGTCTCTCTTGTAGTTATGATG 73
Db	129205 ATTATTGGAGGCGAGTAAATFAAATGAGTAATGAGTAATCACCTGTAGCAATATTATTATCTCTCGATG 129146
QY	74 GATGGGCGCTTCCGATGAACAACACGCGGAATCAATTTGCTAAAGCTTAAACGCGCTATTA 133
Db	129145 GTTTTGTAAACGTGCGAAGAACAGTAGGTAAATGCTAGTCTCAAGCAAAACCAAAAT 129086
QY	134 TGGACAAACTTTTCTCGAAATTTGGCA---AAATTTGGAAGCACACGCTCTTCATGTTG 190
Db	129085 TCGACCGCTTATGGGCTGATTTTCCCTCACGGGAACCTTAAAGCTGCTGGGCTTCATGTTG 129026
QY	191 GATTGCCAGAGGCTTAATCGGAAATTCCTGAAGTTGGACATTTGAATATAGGAGCTGGAA 250
Db	129025 GCCTTCCAGAAGGTCAAATGGGTAACTCTGAAGTTGGCCATACAACATCGGAGCTGGAC 128966
QY	251 GAGTTATTTATCAAGATATTGTTCCGAATTAATTTGGCTGTTTCAACGAAACGAGTTTGT 310
Db	128965 GTATTGCTACCAAGAGCTTAACTCGTAATGCAAAAGCTAATGAAAGGCGAAATTCOAAG 128906
QY	311 CAATCCTCAGATTGTTGCATCAGCTGAGCGTGCAGAAAGGGGAGTGGTCGATTGCATT 370
Db	128905 AGAACAAGCCCTAAACAATGCTTTCACCTATACAAAAGAAACAACCTCGGACTTACATC 128846
QY	371 TATTAGGACTGGTTAGCGATGGTGGTCCACTCTCATATTTGATCATCTTTTGGCGTTGA 430
Db	128845 TTTTTCGTTTACTATCAGAGCGGGCGTGCATAGTCACATTAATCACTCGTGGCACTTT 128786
QY	431 TACGTGCATTAAACRAATTAACAAGTGCCAAAGTTTTCATCTACTTTTTCGTCATGGTC 490
Db	128785 TAGAACTGCGAAGATTAAGGCGTGAAGGCTTTTATATCCATGCAATCTCTTGATGGAC 128726
QY	491 GAGATACTTCCCAACAAGTGGAGCTGGTTATCTTGAACAACCTTCTTCAATTTATTGCTT 550
Db	128725 GTGACGTGGCCACCACTCCTCACTAGAAATATTTAGAAACACTACAAAAGCTATTAGTG 128666
QY	551 CGGAAAGTAGGAGAAATTTGGCTACTATTACTGACGTTTATATGCAATGATAGGAGACA 610
Db	128665 ACTTAAACTATGGCGGATTTGCTACTGTTTCTGGACGTTTCTAGCGATGGATGCTGATA 128606
QY	611 AAAGATGGGCGGTATTAAAGATGCTTATGAGGCAATTTGTTGGAGGTATTGGACAAAGAAG 670
Db	128605 AACCTGGGACGTGTTGAAAAGCATACAAAGCAATCGTAAGCGCTGAAGGTGAAAAG - 128547
QY	671 CCACCGTTGATAAGGCTGTCGATGTTGTTAGAGAGCGATATGCTCAATCTGAGACTGACG 730
Db	128546 -----TTTGAAGATCCCAATCGAACTTCTCAAGCTTCTTATGCTAATGACAAAATCATG 128492
QY	731 AATTCTGAACCAATTTGTTTTCGACGACATGGCC-----GAGTAAAGATG 778
Db	128491 AATTGCTGTCTGCTATCATTTACTAAAGATGGCAAAACCTGTTGCAACAGTTTAAAGACA 128432
QY	779 ACGATACCTTATTTCTTCAATATATGCTGCTGATCTATGCTGCTCAATTTGCTGAATGTT 838
Db	128431 ACGATGACGATTTTCTTCAATTTCCGCTGCTGACCGTGGATTCACCTTCTTAATGCAT 128372
QY	839 TGGCTCTCGAACGTTATAAGATCTTAAATAGTTCGGTTCCTCACCCCTAAAAATATTGACA 898
Db	128371 TCACTGATAAAGATGGATCATTTTCGACCGCTGGGCGAGATCACCCCTAAAAACATAAT 128312
QY	899 TTAGTGGGATGACCCATACAAATAAGAGTTTCCATTTCCATTCGTTATTCACCTGTGA 958
Db	128311 TTGTACAAATGACTCTTTTACAACCCCAAGTATTGATGCTGAAGTTGCTTTTGGACCAATG 128252
QY	959 CTCATATAATGCTGCTGAATGGCTTCTTCTCAAGGAGTTTACTCAATTTTACCTGTCG 1018
Db	128251 AATGAAAATTTTATCGGCGAAGTACTTTTCTAATGAAGGCTTATCGCAACTGCGTATCG 128192
QY	1019 CGGAAACTGAGAAGTATCTCATGTTACCTTCTTCTTAAATGTTGGTGGTCGAGAAAGTTCAAT 1078

Db	128191 CTGAAACAGAAAAATATCCACACGTAACATTTCTTTATGATGGTGGACGAAATGAAGAAT 128132
QY	1079 TCCAAGATGAGAGCGTGTGATGGTCCGTCACCAAAAGAGTTGCTACATATGATTTAA 1138
Db	128131 TCCCTGGTGAACCCGGAATTTCTAATCAATTCGCCAA---AAGTAGAACAATACAGATTGC 128075
QY	1139 AACAGAAATGAATGCTGCTGGAGTTGCCGAAAAAATGGTCGAGCAAAATGAGTCAGCA 1198
Db	128074 AACCTGAATGAGCGCATATGAAGTAACAGATGCACCTGTTGAAGACATTTAAACAGCA 128015
QY	1199 GGCATCCTTTGGTTATGTGCAATTTTGGCTCTGACATGGTTGGACACTACTCGTAAAT 1258
Db	128014 AACAGATGCGATCATCTTAAACTTCGCAAAACCCAGATATGGTTGGCACTCAGGTATGC 127955
QY	1259 TTGAACCTGCGTCAAAAGCATGCAAGCTACTGACGAGGCAATTTGGAAAGATATTTGAAG 1318
Db	127954 TTGAGCCAAAGATTAAGCAATCGAAGCAGTAGATGAAATCTTGGTGGTGTAGTAGACC 127895
QY	1319 CATGCCAAACTTATAATTAAGTTAGCTTCTTATGGTTACTTCCGATCATGGAATGCTGAGAAGA 1378
Db	127894 TTATTTAGAAAAGTGGTTTCAGCTATTATCTTGTGCTGACCATGGTAACTCTGAACAA 127835
QY	1379 TGATTGCTCCCGATGGTAGTGAACATACTGCACATACCTGCAATTTGGTCCCATTTACT 1437
Db	127834 TGCTACTCCAGAAGAAAACCCGACACTGCGCACACTACGGTTCACAGTTCACAGTAAT 127776
RESULT 9	
AP003190/c	
LOCUS	AP003190 301150 bp DNA linear BCT 10-JUL-2002
DEFINITION	Clostridium perfringens str. 13 DNA, complete genome, section 6/10.
ACCESSION	AP003190 BA000016
VERSION	AP003190.2 GI:18146728
KEYWORDS	
SOURCE	Clostridium perfringens str. 13
ORGANISM	Clostridium perfringens str. 13
	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
REFERENCE	
AUTHORS	1 Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A., Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.
TITLE	Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
MEDLINE	21664373
PUBMED	11792842
REFERENCE	2 (bases 1 to 301150)
AUTHORS	Direct Submission
TITLE	Submitted (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical Sciences, University of Tsukuba, Department of Microbiology; 1-1-1 Tennohda, Tsukuba, Ibaraki 305-8575, Japan
JOURNAL	(E-mail:tschimu@md.tsukuba.ac.jp, Tel:81-298-53-3354, Fax:81-298-53-3354)
COMMENT	On Jan 14, 2002 this sequence version replaced gi:18144942.
FEATURES	Location/Qualifiers
source	1..301150
	/organism="Clostridium perfringens str. 13"
	/mol_type="genomic DNA"
	/strain="13"
	/db_xref="taxon:195102"
	/note="anaerobic pathogen for gas gangrene"
gene	complement(3..5066)
	/gene="CPE1281"
CDS	complement(3..5066)
	/gene="CPE1281"
	/note="1687 aa, partially similar to pir:T36472 probable secreted alpha-galactosidase from Streptomyces coelicolor (680 aa); 33.6% identity in 140 aa overlap. Putative N-terminal signal sequence was found by PSORT"
	/codon_start=1
	/transl_table=11

```
/product="hypothetical protein"
/protein_id="BAB0987.1"
/db_xref="GI:18144943"
/translation="MNKRIATAIATMTNLSATTIDVLAQELNTKNSKVEVSHDD
ESHQARVFDLYNSDKLDAYNQEFQINRSKITSINNGKYNSSIDKRAIDGNLETH
WETKPNDSNFNEVVVTEITNDRIYISARDSARGKFAKEIDYIASLKQEDDD
FNIVSGEYESTRDLVEIKFNPDKRLKFKPKADQNWASSAFPMFKEDLNKFK
NGLFTDSMNKVEEETLEKLNAPENELKDPMDLYLKEGLNNARAILTESENPTK
ATLGQITVNLNDYNOYRMPYTNISIKNGRHYAAQNIKRAIDNDVNTYETGTLN
SSGFNEVEENDLVTLDRIVYGSOSDLKGFPAEVIYASRTSGDYTKLVATGAH
EATKGLVEAFEPTEKRVKFKKQNSATINELMFKPDEVISISPKLFTDGTMS
EUSEENSLEKINAFKAKNHPDLYNDFKETIDLAESLISNPKRDEVDLEMRGDIS
ELKRWVNFQWQITGLSARAGDKITVYVDAEGDPTLLYKQSLTGHGATSFOL
KPKNEITPEINYESGIPKDVIOGQDLFFNYKSDSKRAPKVRIEGASKYVPFIL
KSDNEVKELEAYVEKIKAEKPTPNIFAVSSNKSLEFVOATYALDVKNKPTPK
YIAEOWDQYIADAGPWFQDNDKVDNSDFNERIMPVKNLSGAFNAGNVGIRPG
NQDAILAANGWGAHELGHNDTGRTIVEVNNMPLDFFESKYTKKTRIDQNIWE
NNTYKVGDDYSNNELNKADSTHLAQLAPQLYLDLYNTFFYKFERFRERDNK
NREDIYKSWVAASDAEMLDTEFFARHGIRVDKVKEDLAKYPRDKKIYINDLAM
NYKGDGTTNARYSVSTSGSNGIKLSFVDNENKDNILGIEIRDKGYVTGTSNDSF
VOTKSLDEGTVVTPYDKRLNTLNPINEVALQPTLSVNPVITTLALGEFNEEYIV
AKDKISLSESVKVKSNNVTSKGEYEVLYSLEDKNGEYTKTSKVNVSREKYS
DLTPKSSNGWTVRKDKSISGGVIGLTRDGFVDYKGLGHSNAEYVYDLEGDYD
YFESYGVYDKMSSRPASSYIFKVLVDGEEKNSGVRSTTPQKYVYVDVNAKELKL
IYNDAGDSDSHASFGDAKATLSKPIIKGENLAYNDEKVDLMKGITATDIEDGN
ITSKIQIKSDDFVEKSGIFKVYVSYTSDGLTSECSRTIAVTKETQLSDLNKWSAT
IGSVYKRDRAWSGNQIRLLNEDNSVOTFAKGITSESYSEIYVNSEGYDFTWVGID
RIVADKVSVAFAKVVYDGLKAETDMDVDPKRLVVDVRSKEIKLVVDVADNGN
TWDHADMADKPRNAEYDTTELKALEAKKLDLNNYTESSEFALKNAISGEERALL
SRDKETINSALBELKAMSLVKVDLNAVINIPDKYLLASIONLNTKGTIDILGDMYS
LTTLSGVGDDUSGLNENKAKNLTNNDYNEVDLRLPSKLLKLTNLAQEQFIAAGEL
KFSNGVIGDSYVNRGNKNAVARTIRVDKNGNTILEQADKDEFTINTKDLSSGLYGV
HVLFEDEFGSMVYLFNV"
complement(5515..6903)
/gene="CPEI282"
/gene="CPEI282"
complement(5515..6903)
/gene="CPEI282"
/translation="MDISNTLENVEEVKIKKAVPAVLAIMIFALVDNSFKIISPD
AKAFNISATVSNQVTLGIVIGAVVYASISDSISVRTLLVGLIILICVGLSGLFI
FOKSELIIVISRIQSAGLSAETLYVPIFAKFKSEHKYLGSTSSYAIQSQVIGT
ITGGYSTVLSQVLFILPILALVILPILKIPKKNAGNDVIGFLVAVTAATL
MIFVSGFOWGLIGFIVALLALFLIVSKNENAFISIKFPTNRNFTSVLLAVFVYSQ
LAFIFMPELVESLYNRDQISLLIVPSVLAATVAGLSKVSKIGSKQCIQWAIY
GIIVSLIGALFVKVSVLVEFVWILFSSSFALMAYPMIDTLTRIPREKTTAGFY
NCLNLIATISGITYVAIFMETYKLGKNTGLFSDLVAPIQYSNVLIVLAVISLVALY
WTLIGRREKKDI"
7322..8071
/gene="CPEI283"
7322..8071
/gene="CPEI283"
/translation="MRLATDKPTKSKLLYTSYDKLYKMIIDGEPYPLNSKLPSEPEL
AKSFGVSRMTLQALALQDDGLVKNIRKGNFIISPEKCAKRDLELIGNPIYKCHT
EKIODIEFVRELESDYTKOVLEKDSATVACERYKSKVLAYATFISVETVKE
IKIDNYESLELDFLEHKSYEISESSILEIKHSSAGNSLNPYKLVGNQCDLVLETL
YANDISPPILFNKYIIPSOISTFKIARKKE"
complement(8189..9370)
/gene="CPEI284"
complement(8189..9370)
/gene="CPEI284"
/translation="MKYIIGVGLIVILAWIGSSDKKNIKKVPVFLMLIQFVLGF
ILLNTOGVNLYVSTANGFOVLLDCAGDVNFGVGLVNDQSFISVLLPIFISA
LGILQYLKILFPVIRYIGLVSKINGMGLKESYNASAILQGSVFFISVKKELGSL
PKHRYLTICASAMTSVMSIVGMYLVQPRVVTAVLNLFGGFIASVNPYSVA
EDILVVEDEKKTFFELVIGFVAVVAAVLGVFAIVAIINNIIFRSIFGMS
FOEILGVFSPFAFLMGVPLNESIOAASVMAKLVSNFVAMNLSSGTLNLSARSIG
IYVFLISFANFSSIGIISGAVKGLNQEGNVVAREGLKLLIGATVLSMLTATVGLI
I"
complement(9831..10928)
/gene="CPEI285"
complement(9831..10928)
/gene="CPEI285"
/translation="MENTWEKKWEDIEYLKELKRNHNLFAYTEESFNEKIEKLNK
NMFELDYEEIKVEISRVVASLKDAHTSLIFPAKRFIPLKYFYNDGVVINTCKGYE
NLLFKVLAGDMKEEVEELSNIISFENEYFFRAQSKMKYQIAEVLGLLIIDTMD
KIKITLDGGEYEVITCEDLLYTKNRKLPYARNDSNENLWFOILGSGELLYIKNSCRE
QOESIEKIEKNIICLIEKENIEKTVDLRNLDGSDTLFTLIDLYLKNSEKINKEN
LKVITGRETFSALLNAVTFKNTNAKIIEPSSGKPKNCYGEITRLTLPNSKLIVTYS
TRFYKLIEDSVYALPPEVLESTEDYINK"
11256..11708
/gene="CPEI286"
11256..11708
/gene="CPEI286"
/translation="150 aa, similar to gpu:AB004385.4 conserved
hypothetical protein from Vibrio cholerae (170 aa); 43.8%
identity in 144 aa overlap. Putative N-terminal signal
sequence and 1 putative transmembrane region were found by
PSORT"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB0992.1"
/db_xref="GI:18144948"
/translation="MDLVYVAISVGLSVGMMAAPEGVMSVTFGVGWTGYYSGGG
MAKGLFSLCSNATGIYALLIKGSIFFNPNEYSYFLTAPFSFMVVIQAKPKFEFI
SCAFQCFSTFGLNGWQAVIPALLCGNVLVGISQKIGIYCHKVFSKI"
complement(11999..14614)
/gene="CPEI287"
complement(11999..14614)
/gene="CPEI287"
/translation="871 aa, partially similar to p1r:C72285
hypothetical protein from Thermotoga maritima (strain
MSB8) (718 aa); 22.8% identity in 487 aa overlap. 5
putative transmembrane regions were found by PSORT."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB0993.1"
/db_xref="GI:18144949"
/translation="MKILYKMKYTKOENLIFKGNLTIIISCFMIGILFDNVNF
ILKTYQENKSLSVALLIFLFFIILVKELKLNKIKVRIINDIDFLIASLISASILLI
IYIGLDINKYIEVLSIMVSIIGLVIRIYMINNALMKVNEKNTIYTLKDIENKI
SLSDSHTIFLNEEESYDLERNIIEKLYEAIIVNCNPKRRFIIISLEGNWMSGKTTIL
```


gene /strain="QM B1551"
/db_xref="taxon:1404"
34..1569

CDS /gene="pgm"
34..1569

/gene="pgm"
/transl_table=1
/codon_start=1
/product="2,3-bisphosphoglycerate-independent
phosphoglycerate mutase"
/protein_id="AAB26327.1"
/db_xref="GI:4589056"

/translation="MSKPVALLIILDFALRDEKGNVTHAKPNFDRFNEYPHAT
LQAGGAVGLPEQMGNSGVSHLNIAGRIYQSLTRVNVIREGFEQNETLLAAVK
HAKKGTNLHLLGGLSDGVSHLHLYALLRAKSEGLEKVIYHGLDRDVAPOSA
ETYLKELNIEERYGVEIATLSGRVYSNDROKRWERKSVAMVYGGPSVSAEE
CVKDSYNGIYDFEVLPSVITKEDGSPVATIQDEDAVIFYNRPDRRAIOISNIFANED
FRSDFRGEHPRFLVCLTHFSETVDYVAFKPINLNTLGEVLSONNLQRLRAET
EKYPHTVTFMSGREAEFFETRIILDSKPVATYDLPKMSAYEVTDALAELEGDKQ
DAILNFANPDVHSGMLETYKAIETVDECLGKIYDAILAKGSGTATITADHGNAD
VITLEGNPMTHAHTNPVPIVTKQGLREDGLDGLAPMLTLFLDVAOPKEMTKT
LIK"

BASE COUNT 540 a 318 c 363 g 433 t

ORIGIN

Query Match 17.4%; Score 299.6; DB 1; Length 1654;

Best Local Similarity 52.4%; Pred. No. 9.8e-55;
Matches 737; Conservative 0; Mismatches 654; Indels 15; Gaps 3;

QY 42 TCAACAAAAGTCTGCTGTAGTATGATGATGGGCTTCCGATGAACAACAGG 101
DB 39 TAAAAAGCCATAGCATTAATCATCTAGATGCTTTTGCATTACGGCATGAAGATAAG 98
QY 102 GAATGCAATTCGTAAGCTAAACGCCCTATTATGGACAACTTTGTCGGAATGGCA 161
DB 99 TAATCGGTTAACACATGCRAAAAAACCAACTTCGACCGTTCTGGAACGAGTATCCTCA 158
QY 162 ---AAAATGGAAAGCACAGCTTTCATGTTGGATTGCCAGAGGCTTAATGGAAATTC 218
DB 159 TGCTACACTTCAGCGCTCTGGAGAAGCTGTAGCTTACCAGAAGGCCAAATGGTAACTC 218
QY 219 TGAAGTTGGACATTTGAATATAGGAGCTGGAAGAGTATTTATCAAGATATGTTTGAAT 278
DB 219 TGAAGTAGGTCATTAACATCGCGCAGCGCAATTTGTACCAAGCTTAACTCGTGT 278
QY 279 TAATTTGGCTGTTCACAGAAACAGTTTGTACAAATCCTCAGATTTGTCATCAGCTGA 338
DB 279 AAACGTAGCAATTCGTGAAGCGAGTTGAACAAACGAAACGCTTCTAGCAGCTGAAA 338
QY 339 GCGTGAAGAAGAGGGAGTGGTCGATTCGATTTATAGGACTGGTTAGCGATGGTGGTGT 398
DB 339 GCACGCAAAAGAAAGGTAACAAACCTTCATCTTTTCGGGCTTTTATCTGATGGCGGTGT 398
QY 399 CCACTCTCATATGATCACTTTTTCGGTTGATACGTGCTATTAACAATTAACAAGTGC 458
DB 399 ACACAGTCATATGAACACTATATCATATTGCGTCTAGCTAGCTAGAGGTGAAGCTTAGA 458
QY 459 AAAGGTTTCACTTCACTTTTTCGTCGATGCTGAGATACCTTCGCCAACAAAGTGGAGCTGG 518
DB 459 AAAGTCTACATTCATCGCTTCTTAGTGTGCTGACGTAGCACCTCAATCTGCTGAAC 518
QY 519 TTATCTTGACAACTTCTTCAATTTATGCTTCGGAAGAAAGTACGGAGAGATTTGGCTACTAT 578
DB 519 GTATTTAAAGAACTTAATGAAAAAATGAAGAGTACGGGCTTGGTGAATTTGCAACACT 578
QY 579 TACTGACGCTTATATGCAATGATAGGACAAAGATGGGACGCTATTAAGATGGCTTA 638
DB 579 TACAGACGCTTACTACTCAATGACCGCAGCAACACGCTGGAGCGGTAGAGAAATCATYA 638
QY 639 TGAGGCAATGTTGGAGGTATTGGACAAAAAGCCAGCTTGTATGAAGGCTGTGCTATGCT 698
DB 639 TCGCGCAATGGTACGCGGAGGTTCATCATATACAAGCGCTGAAGAGTGGTAAAGA 698

QY 699 TAGAGAGCATATGCTCAATCTGAGACTGACGAAT-----TTCTGAAACAATTTGT 749
DB 699 TTGCTATGACAATGGAATCTATACGAATTCGTTTACCTTCTGTTATACAAAGAAGA 758
QY 750 TTTTTCGGACGATGGCGAGTAAAGATGACGACTACTCTATTTTCTTGAATTTATCGTGC 809
DB 759 CGGATCGCGAGTTGCGACAATTCAGCAAGAGATGAGTATTTTCTACAACTTCGCTCC 818
QY 810 TGATGCTATGCGTCAAAATTTTGTGAATGTTTGGGCTCTCGAAGCTTATAAGATCTTAAATAG 869
DB 819 GGATCGCGAATTCAAATTTTCAATAGCTTCGCAAAAGAGATTTCCGTTTCATTTGATCG 878
QY 870 TTCGGTTCTCACCCCTAAATAATATCAGATTTAGTGGGATGACCAATACAAATAAGAGTT 929
DB 879 CGGTGAAAAACATCCAAAACACTTACATTTTGTATGTTTAACTCACTTCACGCGAAACAGT 938
QY 930 TCCATTTCCATCGTTATTTCCACCTGTGACTCATATGCTTGTGCTTGTGCTGAAATGGCTTC 989
DB 939 TGACGGATATGATGACCTTTTAAAGCAATTAACTTGGATAACACGCTTGGTGAAGTATATC 998
QY 990 TTTCTAAGGAGTTTACTCAATTTTCACTGTGCGGAAACTGAGAAGTATCTCTCATGTTACCTT 1049
DB 999 TCNAATAACTTAACAGCAGCTTCGTATTGCCGAGACAGAAAATATCTTCACGTAAGCTT 1058
QY 1050 CTCTTTAATGTTGGTGGAGAAATTCATTTCCAAAGTGAAGACGCTTGTATGTTCCGCT 1109
DB 1059 CTTTATGCGGTGGAGCTGAAGCAGAGTTTCTGCTGGAACACAGCTATCTTAATCGATTC 1118
QY 1110 ACCAAAGAAGTTGCTACATATGATTTAAACCAGAAATGAATGCTGCTGAGTTCGCCGA 1169
DB 1119 ACCAA---AAGTAGCAACATATGACTTGAACCTTGAGTGAAGTGAAGTGAAGTGAAGTGA 1175
QY 1170 AAAAATGGTCGAGCAAAATTTGAGTCAGGAGGAGCTCTTTGTTTATGTCGAAATTTTCGCC 1229
DB 1176 CGGTTGCTGCAAGAAATCGAGGCGATAGCAGAGCTATTTTATTAACCTTTGCATA 1235
QY 1230 TCTGACATGTTGGACATACCTGTAATTTTGAACCTGCGCTCAAAAGCATGTCAAGCTAC 1289
DB 1236 TCTGATATGTTAGGCACTTCAGTATGTTAGAACCACCTGTAAAAGCGATTCGAAACAGT 1295
QY 1290 TGACGAGGCAATTTGGAAGATATTTGAAGCATGCCAACTTATAATACGTTCTTATGTT 1349
DB 1296 AGATGAGTGCTTAGGCAAAATTTGATGCAATTTTAGCTTAAAGCGGTACACAACTCAT 1355
QY 1350 TACTTCGATTCGAAATGCTGAGAAGATGATGCTCCGATGCTAGTGAACATACGCTAC 1409
DB 1356 CACAGCAGACCATGGTAAATGCTGATGAGTGAAGTATACGCTTGAAGGTAATCCATGACGC 1415
QY 1410 ACATACCTGCAATTTGGTCCCAATTA 1435
DB 1416 TCATACACGAATCCTGTTCCAGTAA 1441

RESULT 11

AE017014/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AE017014 300854 bp DNA linear BCT 16-MAY-2003
Bacillus cereus AFCC 14579 section 17 of 18 of the complete genome.

AE017014 AE016877

AE017014.1 GI:29898525

Bacillus cereus AFCC 14579

Bacillus cereus AFCC 14579

Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus; Bacillus

cereus group.

1 (bases 1 to 300854)

Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B.,

Kapralov,V., Bhattacharyya,A., Reznik,G., Mikhailova,N.,

Lapidus,A., Chu,L., Mazur,M., Goltsman,E., Larsen,N., D'Souza,M.,

Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fongstein,M.,

Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.

Genome sequence of Bacillus cereus and comparative analysis with

Bacillus anthracis

Nature 423 (6935), 87-91 (2003)

QWYIAKTLNLEASGVVFVVALSVPRLEIFMIVGAVADFPKKNIMFYNIIRAILVA
TITWFIQDITLYTALFALFGLADAFWSADGSILPELVEKSLTOANSITOMTN
QASVILGVGLILKFTNVEIFISITILLVAAILVOKIOFTPEOKDOKGNETS
IKEGILYVESPELSFLCSAFLNLLFGPMQVGPPLFKVNLHGDSDLOFYLEASV
GGMALGVAGLKNLNRRLGFCIIMLLSGVFFLSINFSTVLMQALLAGMFYGITI
ANAILPMIAOSTVKEEMMGWRVMSLLMLSSMGFIPLSYAFSTIALAIGIPIVTMKS
GAIAIVFVLVAIRVPVVRKFD*
complement(15379..16266)
/locus_tag="BC4930"
complement(15379..16266)
/locus_tag="BC4930"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator, Deor family"
/protein_id="AAP11803.1"
/db_xref="GI:29898531"
/translation="MGSLYNLAKPKYSEFQSKSEFNYYQKVLKHYRFQNLKNDATIIH
FGKYAVNKKQTVGACPLMETIATNVKRSINTVRRSIAKLEELGIIIRKVAIKENHR
RGYISANLYVLISADIDRMDRMKSACSEDTAGCSKEQKSERETILSKNIPOIK
EKRTIYLEDYTCRHDIKPIFYALLPMTSNPKINIFWSKVELAYKKSGLLEQVL
LEQILADEEVYGNFIMRVKSVVRVAYKGEIRKNVRALLYSTVRDLFLEVLEWGAVR
RSKGIPLFPFKKEPCVNA*
16347..16808
/locus_tag="BC4931"
16347..16808
/locus_tag="BC4931"
/codon_start=1
/transl_table=11
/product="hypothetical Cytosolic Protein"
/protein_id="AAP11804.1"
/db_xref="GI:29898532"
/translation="MIVISACLGACIACRYDGNLNVLSKEELLOKEDTVLICEVLGG
LTPPSPABIIIGDNGDVLDDKAKVMTKDEVDTEAFVSGYKALEQIKNLHPEYIIL
KERSPSCSGSTIYGFENGNKOTGFGVTTLFALPKRHFVKYVISEEFENKRN"
complement(16839..17192)
/locus_tag="BC4932"

Query Match 17.4%; Score 299; DB 1; Length 300854;
Best Local Similarity 52.6%; Pred. No. 5.6e-55;
Matches 728; Conservative 0; Mismatches 640; Indels 15; Gaps 3;
QY 65 TTATTCATGATGGGGCCCTTCGGATGAACACACACGGGATCGAATCGAATGCTAAAGCTAA 124
DB 205559 TTCCTGACCGTTTCGGACCTCGTGAAGAACTTACGGGAATGCTGTAGCAGCTAAGA 205500
QY 125 CGCCTATTATGACAAACTTTGCTCTGGAATTTGSCA---AAAATTTGGAAGCACACGGCTC 181
DB 205499 AACCTAATTTTGATGTTTGAACAAATTCCTCCACACACGGCTTACAGCTTTGTGGCG 205440
QY 182 TTCATGTTGGATGCCAGAGCTTAATGGGAAATCTGAAGTTGGACATTTGAATATAG 241
DB 205439 AGGCAGTAGGCTTCCAGAAAGTCAAAATGGGTAACTCTGAGGTTGGTCACTTTAAATATCG 205380
QY 242 GAGCTGGAAGAGTATTATTAAGATATTGTTGGAATTAATTTGGCTGTTCAACGAAAGC 301
DB 205379 GTGCTGGCGCATGTATATCAAGCTTAACAGCGGTTAACCTTGCAATTCGTGAAGGTG 205320
QY 302 AGTTTGTTCACAACTTCAGATGTTGTGCATCAGCTGAGCGTCAAGAGAGGGAGGTGGTC 361
DB 205319 AGTTTCATAGAAGCAAGCTTCCAAAGTCAATTAAGACGCTAAAGAAAAAGGTACTG 205260
QY 362 GATTCGATTTATAGGACTGGTTAGCGATGGTGGTGTCCACTCTCATATTGATCATCTTT 421
DB 205259 CACTTCATTTATTCGGTTTACTTTCTGAGGGTGGTGTGCACAGTCACATGAACCATATGT 205200
QY 422 TTGCGTTGATACCTGCATTTAAACAAATTAACAAGTGCACAAAGGTTTTTCATTCACCTTTT 481
DB 205199 TTGCTCTTCTTCGCTTAGCAGCTTAAGAGAGCGTTGAGAAAGTATACATTCATCATCT 205140
QY 482 CTGATGGTGGAGATCTTCGCCAACAAAGTGGAGCTGGTTATCTTGAACAACTCTTCAAT 541
DB 205139 TAGATGGCGCGCATGTGTGACCAACAAACAGCAAGCTATATCGATGCAACAAATGAAG 205080

QY 542 TTATTCCTTCGGAAAAAGTACGGAGAAATTCGGCTACTATTACTGACGCTTATTATGCAATGG 601
DB 205079 TAATTAAGAAACAGAGTAGGACAATTCGGAGCATCTCTGCTGCTTATTACTCCATGG 205020
QY 602 ATAGGACAAAAGATGGGAGCGTATTAAAGATGCTTTATGAGGCAATGTTGGAGATTTG 661
DB 205019 ACCGTACAAAGCGTTGGATCGCTAGAAAAATGTTTACCGTGTCTATGGTGAATGGTGAAG 204960
QY 662 GACAAAAACCCACCGTTGATAGGCTGTGATGCTTCTTTAGAGAGCGATATGCTCAATCTG 721
DB 204959 GCCTACTTATAATCAGCAGAGAGAGTGTGTAAGACTCTTATGCGAAGCGTATCTATG 204900
QY 722 AGACTGACGAATTTCTGAAACCAATTTGTTTTTCGGACGATGGCGAG-----TAA 772
DB 204899 ATGAATTCGATTTCGCTCTGTAATTTTAAACGAAGATAACACGCCAGTTGCAACAATCA 204840
QY 773 AGATGACACATCTCTTATTCTTCAATTAATGCTGCTGCTGATGCTGCGTCAAAATTTG 832
DB 204839 ATGATGATGATGACGATTATCTTCTAATACCTCCGTCAGACCGTGCAATTCAAATTTGCTC 204780
QY 833 AATGTTTGGGTCTCGAACGCTTATAAAGATCTTAAATAGTTTCGGTTCCTCACCCCTAAAAATA 892
DB 204779 GTGATTTACAACGAAGACTTCCGTCGATGCTGCTGTAAGAAAGTACCTCACATTC 204720
QY 893 TTCAGATTAGTGGGATGACCCCAATACAATAAAGAGTTTCCATTCATTCCTCCAC 952
DB 204719 CAGAAATTCGATGTATGACACACTTCAGTGAAACTGTAGATGTTACGTTGCGATTCAAGC 204660
QY 953 CTGTGACTCATACTAATGTGCTTGTGAATGCTGCTTCTCAAGAGTACTCAATTTTC 1012
DB 204659 CAATGAACCTTGTATACACTTTTAGGTGAAGTTGTTGCACAACGGGATTAAGCAACTTC 204600
QY 1013 ACTGTGCGGAAACTGAGAGATATCTCATGTTTACCTTCTTCTTAAATGTTGTCGAGAAG 1072
DB 204599 GCATCGGGAACCTGAAAGATATCCGACGTTACATCTTCTTTACCGGTGCTGAGG 204540
QY 1073 TTCAATTCAGATGAAGAGCGTTGTATGGTTCCGTCACCAAAAGAGTTGCTACATATG 1132
DB 204539 CTGAATTCACGAGGAAGAGCGTATCTTAATTAACCTCACCAA---AGTTTGAACGATG 204483
QY 1133 ATTTAAACCAACAAATGATGCTGCTGGAGTTGCCGAAAAAATGTCGAGCAATTTGAT 1192
DB 204482 ACTTGAACCTGAAATGAGCATTTACGAAGTAAACGACGCTTTAGTAAATGAATCGAAA 204423
QY 1193 CAGGACGAGCTCCTTTGGTTATGTGCAATTTTTCGCGCTCCTGACATGTTGGACATCTG 1252
DB 204422 ATGATAAATCATGATGTTATCTTAACTTTTTCGGAACGTGATATGTTGGCCATTTCTG 204363
QY 1253 GTAAATTTGAACCTGCCGTCAAGCATGTCAGCTTACTGACGAGGCAATTTGGAAGATAT 1312
DB 204362 GGATGATGGAACCAACAATTAAGCAGTAGAAGCAACTGACGAATGTTTAGAAAAAGTTG 204303
QY 1313 TTGAAGCATGCCAAACTTATAATTTACGTTCTTTATGGTTACTTCCGATCATGAAATGCTG 1372
DB 204302 TAGAAGCATTTCTTCAAAAGATGTTGATGACACTTATTACTGCTGACCATGTTATGCGG 204243
QY 1373 AGAAGATGATTGCTCCCGATGTTAGTGAACATPACTGCACATACCTGCAATTTGGTCCCAT 1432
DB 204242 ATGAAGAATTAACCTTCTGAAGGAGAGCAAAATGACAGCTCATACAACCTAACCGGTTCC 204183
QY 1433 TTA 1435
DB 204182 TCA 204180
RESULT 12
AE017040/c
LOCUS AE017040
DEFINITION Bacillus anthracis str. Ames section 17 of 18 of the complete genome.
ACCESSION AE017040
VERSION AE017040.1
KEYWORDS GI:30259630

AE017040 Bacillus anthracis str. Ames section 17 of 18 of the complete genome.

ACCESSION AE017040
VERSION AE017040.1
KEYWORDS GI:30259630

SOURCE	Bacillus anthracis str. Ames	
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.	
REFERENCE	1 (bases 1 to 291205)	
AUTHORS	Read, T., Peterson, S., Tourasse, N., Baillie, L., Paulsen, I., Nelson, K., Tettelin, H., Fouts, D., Eisen, J., Gill, S., Holtzapple, E., Okstad, O., Helgason, E., Rillstone, J., Wu, M., Kolonay, J., Beanan, M., Dodson, R., Brinkac, L., Gwinn, M., DeBoy, R., Madupu, R., Daugherty, S., Durkin, A., Haft, D., Nelson, W., Peterson, J., Pop, M., Khouri, H., Radune, D., Benton, J., Mamoud, Y., Jiang, L., Hance, I., Weidman, J., Berry, K., Plaut, R., Wolf, A., Watkins, K., Niernan, W., Hazen, A., Cline, R., Redmond, C., Thwaite, J., White, O., Salzberg, S., Thomason, B., Friedlander, A., Koehler, T., Hanna, P., Kolsto, A.-B. and Fraser, C.	
TITLE	The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria	
JOURNAL	Nature 423 (6935), 81-86 (2003)	
MEDLINE	22608414	
PUBLISHED	12721629	
REFERENCE	2 (bases 1 to 291205)	
AUTHORS	Read, T., Peterson, S., Tourasse, N., Baillie, L., Paulsen, I., Nelson, K., Tettelin, H., Fouts, D., Eisen, J., Gill, S., Holtzapple, E., Okstad, O., Helgason, E., Rillstone, J., Wu, M., Kolonay, J., Beanan, M., Dodson, R., Brinkac, L., Gwinn, M., DeBoy, R., Madupu, R., Daugherty, S., Durkin, A., Haft, D., Nelson, W., Peterson, J., Pop, M., Khouri, H., Radune, D., Benton, J., Mamoud, Y., Jiang, L., Hance, I., Weidman, J., Berry, K., Plaut, R., Wolf, A., Watkins, K., Niernan, W., Hazen, A., Cline, R., Redmond, C., Thwaite, J., White, O., Salzberg, S., Thomason, B., Friedlander, A., Koehler, T., Hanna, P., Kolsto, A.-B. and Fraser, C.	
TITLE	Direct Submission	
JOURNAL	Submitted (26-VAR-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA	
FEATURES	Location/Qualifiers	
source	Join(1..170086,186846..291205)	
	/organism="Bacillus anthracis str. Ames"	
	/mol_type="genomic DNA"	
	/strain="Ames"	
	/db_xref="taxon:198094"	
	/focus	
source	complement(170087..186845)	
	/organism="Bacillus anthracis phage lambda Ba02"	
	/proviral	
	/mol_type="genomic DNA"	
	/db_xref="taxon:229344"	
	/note="Putative prophage; Putative lysogenic prophage region. Best Hits not completed. Putative attL/R"	
gene	90..803	
	/locus_tag="BA5145"	
CDS	90..803	
	/locus_tag="BA5145"	
	/note="similar to GP:4584089, and GP:4584089; identified by sequence similarity; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAP28817.1"	
	/db_xref="GI:30259631"	
	/translation="MHFFVQALQEHFTANQNPKEAPMAYMKNHFFLGIQTPERRQ LKDVQIHTLPDPKDFQIIVRELWDLPERFOAALDMQYKKNHETHIFPLEEL IVTKSWMDVDSIVPTFLGNFLKHPLEISATPKWTASDNIWLQRAAILFOLKYQK MDELLFWIGQLSHSKEFFIQKAGWLVREYAKTRKPDVVVEYQNNELAPLSKREAI KHKENYGINEXIGETLS"	
gene	974..1105	
	/locus_tag="BA5146"	
CDS	974..1105	
	/locus_tag="BA5146"	
	/note="identified by Glimmer2; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AAP28818.1"	
	/db_xref="GI:30259632"	
	/translation="MDAPFYQLGWNRRGQTLVPRQLSLGMSVFLCFCKSHLLYSQVT" 1151..2527	
	/gene="glsY"	
	/locus_tag="BA5147"	
CDS	1151..2527	
	/locus_tag="BA5147"	
	/note="similar to SP:P75425; identified by sequence similarity; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="glycyl-tRNA synthetase"	
	/protein_id="AAP28819.1"	
	/db_xref="GI:30259633"	
	/translation="MYSMEQVNLAKHGFVPCSEIYVGLANTWDVGPGLGIELKNV EDALDKAGIEVSYNVLGDLDAAILMNPNTIASGVGNFNDPMDICKCKARHADKLI KRAWKRFIOESPNVGLDQADLMDLMEHYKPCDCSEETEFIRQFNLMPKTFQGVTE SQMTEIFLRPETAGGIFVNFKNVQSRMRKLLPFGIGQIGKSRNEITPGNFTFTREF EOMELEFFCKPGEGLWFVAFWRETCKNLLSLQFTEESMRLEDRHGEELSHYSNATFD IEFKFPFGWELGAVASRTDFDLKRHMEHSENFNYIDPQTNERYVPCIEPISLGAADR VTLAFLCDAYEEQLENDSRTVLRHFPALAPYKAAIPLSKLSEGADEVFAELAKDF MVDFTGSGIKGRYRRODEIGTFCITYDFDSVEDKAVTVDRDRDTRMEQVRMPISELKG FLEKKIQF"	
	complement(2565..2948)	
	/locus_tag="BA5148"	
CDS	complement(2565..2948)	
	/locus_tag="BA5148"	
	/note="similar to GP:10119873; identified by sequence similarity; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="comA operon protein, putative"	
	/protein_id="AAP28820.1"	
	/db_xref="GI:30259634"	
	/translation="MAKTLMDSLGIELLEMTEDKVATMPVDERTHPQFGLHGGASL ALAETVASGVSYNLIDQECICFGLSEINANHIRSKDGMVTAIGTPHRRGHSTMWDV RIIDNDLLICISRCTVAIKEKREK"	
gene	3044..3787	
	/locus_tag="BA5149"	
CDS	3044..3787	
	/locus_tag="BA5149"	
	/note="similar to SP:P19196, GB:X53368, GB:Z48169, PID:48574, and PID:793895; identified by sequence similarity; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="conserved domain protein"	
	/protein_id="AAP28821.1"	
	/db_xref="GI:30259635"	
	/translation="MKOKFFILSLMFLSLITVPCALAAKASNLITDVQTVTKGKKP YIEQINRPSFHNFSDFKFNKFNFKYKSTDKPKTELEKDAKYYKKEGSSIPFHP DYKVDKSLNRPPLSLXUNTYQYGGAGHLYTWKANTFDLKEKLLHLDLDFQOE DKYKEVIRAEIRKIKONEGIYFPDATEKVMSTKKKFFHLEPNLNLVIYFPLYEIAPYS SGIPQFRIPYTLRLREYLKPSYQNILDNK"	
gene	3838..4431	
	/locus_tag="BA5150"	
CDS	3838..4431	
	/locus_tag="BA5150"	
	/note="similar to GP:10175981, and SP:P04028; identified by sequence similarity; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="biol family protein"	
	/protein_id="AAP28822.1"	
	/db_xref="GI:30259636"	
	/translation="MRRFHVLDLDAAMFVALMAIGANIVSWAPFLQVAGIPLSMQPF FAIAGLLGLSRGALSMVTVMVLGVAGAPIFANFKAGFGALLDGTGTFIATIVAY VSKLVEQKEKPTFTSTAIASFTGIIITYIGTYMYGAVNLFMNGNMSYKAAWMIMM WFAVKDIVFTIIGAIIPRIYVAVRRSAYQHSHTIS"	
gene	4579..5415	
	/locus_tag="BA5151"	
CDS	4579..5415	

```
/locus_tag="BA5151"  
/note="identified by match to PFAM protein family HMM  
PF02978"  
/codon_start=1  
/transl_table=1  
/product="hypothetical protein"  
/protein_id="AAP28823.1"  
/db_xref="GI:30259637"  
/translation="MNLKNTLIVGSLVGLFFILGMSIGYKLVGNSLGMSTYFGN  
GLAFLIITIVSLCAALFIVIGLKKVKKLPKAKSVATSIILFVTSIISLIVLFT  
IFKVNMEEXYQVLQOKKKKEASLYLVAASFYNNINFNAYASYLVSESTWTSNAD  
KQDFNNALSDIEDGMITVDYFISTGNDLKLVSAAKEQPNKYKETEEYKKI  
YGIITLNEQASPSGSLISFNQNVNLIQYKKAAGNINIAITDEIKSKANELKPTD  
KN"  
gene  
complement(5521..6408)  
/gene="gtaB"  
/locus_tag="BA5152"  
CDS  
complement(5521..6408)  
/gene="gtaB"  
/locus_tag="BA5152"  
/note="identified by match to TIGR protein family HMM  
TIGR00453"  
/codon_start=1  
/transl_table=1  
/product="Urp-glucose-1-phosphate uridylyltransferase"  
/protein_id="AAP28824.1"  
/db_xref="GI:30259638"  
/translation="MKRVKKAIPAGLGTRELPATKAMPKEMLPVDPKPTIQYIVEE  
AVKSGIEDIIIVGTGKTSIEDHFNLEQNLLEKKYELLEKVOASSKVDHYI  
RKEPKGLGHAVWCARKFIDGPPFAVLGDDIVQAEKPCRLQRIIEYDLSVIGVQ  
TYPEDETHYIIDPLEQGRYQVRNFVEKPAAGTAPSLAIMGRVLYLPTFIMFLE  
QDRVAGGSGIQTDAISLNEIQRFVAYDFEGKRDVGEKLGFTVQTTIEMALQHPFLR  
DMVVTMMKKILIEQTNQES"  
complement(6517..8241)  
/locus_tag="BA5153"  
CDS  
complement(6517..8241)  
/locus_tag="BA5153"  
/note="identified by match to PFAM protein family HMM  
PF02878"  
/codon_start=1  
/transl_table=1  
/product="phosphoglucomutase/phosphomannomutase family  
protein"  
/protein_id="AAP28825.1"  
/db_xref="GI:30259639"  
/translation="MNWQEFSRWLSYABDLKLEQLENMKQDEKKIBDSFYKNLEF  
GTGMRGLGAGTNRNLNVYTVRKATGLAKFIEKLGEAEKRGVVYVAYDSRHSPEFA  
MEVAATLGAHGIITVVFESLRPTPLVLSFAVRLHIVSGIVLTASHNPPEYNYKYVGE  
Query Match 17.0%; Score 292.6; DB 1; Length 291205;  
Best Local Similarity 52.3%; Pred. No. 1.4e-53;  
Matches 724; Conservative 0; Mismatches 644; Indels 15; Gaps 3;  
QY 65 TTATTGATGATGGGCCCTTCCTCGATGACACACACCGGAATGCAATTCGTAAGCTAA 124  
DB 189570 TTCTTGACCGTTTCGGAGTCTCGTGAAGAACTTACGGGAATGCTGAGCAAGCTAAGA 189511  
QY 125 CGCCTATTATGACAAACTTTCTCTGGAATTTGGCA---AAAATTTGGAAGCACACGGTTC 181  
DB 189510 AACCTAAATTTTGATGGTTACTGGAACAAATTCCTCCACACACGCTTACAGCTTGGCG 189451  
QY 182 TTCATGTTGGATTGCCAGAGGCTTAATGGGAATTCCTGAAGTTGACATTTGAAATATAG 241  
DB 189450 AGGCAGTAGTCTTCCAGAGGTCAAATGGGCAACTCTCAGGTGTGCTACTTAAATATCG 189391  
QY 242 GAGCTGGAAGACTTATTATCAAGATATGTTGCAATTAATTTGGCTGTTCAACGAAGC 301  
DB 189390 GTGCTGGCGCGCATGTATTAAGAGCTTAACACGGGTAAACGTTGCAATTCGTGAAGGTG 189331  
QY 302 AGTTTGTTACAAATCCTCAGATGTTGTCATCAGCTGAGCGTCAAGAGAGGGAGTGGTC 361  
DB 189330 AGTTTCGATGAACGAACGTTCCAAAGTGCATTAAGAGCGTAAAGAAAGAGGTACTG 189271
```

```
QY 362 GATTGCAATTTATTAGGACTGGTTAGCGATGGTGGTGTCCACTCTCATATTTGATCATCTTT 421  
DB 189270 CACTTCATTTATTCGGTTTACTTTCTGACGGTGGTGCACAGTGCACATGACATGATGT 189211  
QY 422 TTGCGCTTGATACGTGCAATTAACAATTAACAAGTGCACAAAGGTTTTTCATTCACCTTTTG 481  
DB 189210 TTGCTCTTCTTCGGCTTAGCAGCTAAAGAAAGGCGTTGAGAAAGTATACATTCATCATCT 189151  
QY 482 CTGATGCTCGAGTACTTCCGCCAACAAAGTGGAGCTGGTTATCTTTGTAACAACTTCTTCAAT 541  
DB 189150 TAGATGGCCGCGATGTTGGACCCAAAACAGCACAAAGCTATATCGATGAACAAATGAAG 189091  
QY 542 TTATTGCTTCGAAAAGTAGTCGAGAAATTCGCTACTATTAATCTGACGCTTATTAATGCAATGG 601  
DB 189090 TAAATAAAGAACACAGGAGTAGGACAATTCGCGACTATCTCTGCTGCTTATTACTCCATGG 189031  
QY 602 ATAGGACAAAAGATGGGAGCGTATTAAGATGCTTATGAGCAATTTGTTGGAGGATATG 661  
DB 189030 ACCGTGACAAAGCGTTGGGATCGGTAGAAAATGTTACCGTGTCTATGGTGAATGGTGAAG 188971  
QY 662 GACAAAAGCCACCGTTGATAAGGCTGTGATGCTTGTAGAGAGCGATATGCTCAATCTG 721  
DB 188970 GCCTTACTTATAATCAGCAGAGAGAGTGTAGAACACTTATGCAATGGTATCTACG 188911  
QY 722 AGACTGACGAATTTCTGAAACCAATTTGTTTTTCGGACGATGGCGAG:-----TAA 772  
DB 188910 ATGAATTCGTATTCGCGTCTGTAATTTTAAACGAAGATAACACGCCAGTTGCAACAATCA 188851  
QY 773 AGATGACCACTACTCTTATTTCTTCATTAATATCGTCTGATCGTATGCGTCAAAATTTGTC 832  
DB 188850 ATGATGATGATCAGTATCTTCTATAACTTCGCTCCAGACCGTGCATTAATTCAAATTTGCTC 188791  
QY 833 AATGTTTGGTCTCGAACCTTATAAGATCTTAATAGTTTCGGTTCCTCACCTCAAAAATA 892  
DB 188790 GTGATTTACAAACGGAGACTTCGCGTAGTGTGATCGTGGTGAAGAAAGTACCTCACATTC 188731  
QY 893 TTCAGATTAGTGGGATGACCCCAATACAATAAGAGATTTCCATTTCCATCGTTATTCCTCCAC 952  
DB 188730 CAGAAATTCGTATGATGACACACTTTTAGTGAACACAGTGGTGGTGGCATTCGCTCAAGC 188671  
QY 953 CTGTGACTCATACTAATGCTTGTGCTGAATGGCTGCTTCTCAAGGAGTTACTCAATTTTC 1012  
DB 188670 CAATGAACCTTGATACACTTTTAGGTGAAGTGTTCGCGAAGCGGATTAAGCAACTTC 188611  
QY 1013 ACTGTGCGGAAACTGAGAAAGTATCCTCATCTGTACCTTCTTCTTTAATGTTGGTGCAGAA 1072  
DB 188610 GCATCGGAAACTGAAATAATATCCGACGTTACATTTCTTTACGCGTGGTGCAGG 188551  
QY 1073 TTCAATTCAGATGAAGAGCGTTGATGTTCCGTCACCAAAAGAGTTGCTACATATG 1132  
DB 188550 CTGAATTCGAGGAGAGAGCGTATCTTAATTAACCTCACCGA---AGGTTGCAACGATG 188494  
QY 1133 ATTTAAACCCAAATGAATGCTGCTGGAGTTGCGGAAATTTGCGGAAATTTGCTGAGCAATTTGAGT 1192  
DB 188493 ACTTGAACCTGAAATGAGCATTTAGAAAGTAAACGCGCTTTAGTAAATGAATCGAA 188434  
QY 1193 CAGGAGGACTCCTTTGGTTATGCAATTTTTCGCGCTCCTGCACATGGTTGACACTACTG 1252  
DB 188433 ATGATAAATCATGATGTTATCTTCTTAACCTTTTCGGAACGTGATATGTTGGCCATCTG 188374  
QY 1253 GTAATTTGAACCTCGCGTCAAAAGCATGCAAGCTACTGACGAGGCAATTTGGAAGATAT 1312  
DB 188373 GGATGATGAACCAACAATTAAGCAGTAGAAGCAACTGACGAAATGTTAGGAAAAGTTG 188314  
QY 1313 TTGAAGCATGCCAACTTATATTACGTTCTTTTGGTTACTTCGCGATGATGGAATGCTG 1372  
DB 188313 TAGAAGCGATTTGCAAAAGATGTTAGTACTTATTACTGCTGACCATGGAATGCTG 188254  
QY 1373 AGAAGATGATTCCTCCCGATGTTAGTGAACATCTGACATACCTGCAATTTGGTCCCAT 1432  
DB 188253 ATGAAGATTAATCTTCTGAAGGAGAACCAATGACAGCTCATACACTAACCCGGTTCCT 188194  
QY 1433 TTA 1435
```

```
Db 188193 TCA 188191

RESULT 13
ASPLAS
LOCUS
DEFINITION Antithamnion spec plastid DNA sequence.
ACCESSION X64705
VERSION X64705.1 GI:288995
KEYWORDS ATPase; orf 179; orf 510; secA gene; secretory protein.
SOURCE Chloroplast Antithamnion sp.
ORGANISM Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Ceramiales;
Antithamnion.
REFERENCE
1 (bases 1 to 5111)
Valentin, K.
SecA is plastid-encoded in a red alga: implications for the
evolution of plastid genomes and the thylakoid protein import
apparatus
Mol. Gen. Genet. 236 (2-3), 245-250 (1993)
93173098
8437571
REFERENCE
2 (bases 1 to 5111)
Valentin, K. U.
Direct Submission
Submitted (28-FEB-1992) K. U. Valentin, Dept of Botany KB-15,
University of Washington, Seattle, WA 98195, USA
LOCATION/Qualifiers
1. 5111
/organism="Antithamnion sp."
/organelle="plastid:chloroplast"
/mol_type="genomic DNA"
/db_xref="taxon:2767"
203..1735
/notes="orf 510"
/codon_start=1
/transl_table=11
/protein_id="CAA45959.1"
/db_xref="GI:288996"
/db_xref="SWISS-PROT:Q06464"
/transl_table="MNNKSIPIILMDGWHSTNAKNAIHEANTPIIDKLWDNYP
KTLASGEVGHFPHKQMGSEVHTTIGAGRIINODLVKISIKNEKPFNNYQLHN
IYKSEKSNKHVHGLCSNGVSHINHLKALIOISKVKMTCLHLIDGDTLPK
OAKLFVKEIIDIENYNTICITSGRYVSMRDRCWSRTEKAYKAWVENSQNSSS
DLSLIDKYEQNISDEFLLPTKISHNIVHNSDLFFNFRPRRIGOLLSPKAPNPK
GFRKTKLMLTFTFYDSTLSIPTVFPNPKRFLGQIISNSGLKQLRLAETEKIA
HYVFFNGGIEEPFGNRELIASPKVETVDLPKMSAYQLTNSLSEAINKQLYQIV
INYPDMIGHTGNMNTAIEAIEVDQIEKVLHTIEDTNILITSDHGNADYMLTD
ENKPTSHSNPVPFILLNNRQKQTNLHSHGSLADIAP*ILDILNINIPNENKSL
IT"
1759..2298
/notes="orf 179 is a putative homologue to orf X in E. coli"
/codon_start=1
/transl_table=11
/product="non-essential secretory protein"
/protein_id="CAA45960.1"
/db_xref="GI:288997"
/db_xref="SWISS-PROT:Q06463"
/transl_table="MNIQLKNPILKINTQSSNIKLNKIIPNEWLFIIMNTGSFTQ
NLNSLLSDINISQKYNQPOFHNTVWLNANNQRFAPAFASRWIKNNYNY
LNLNNKPGKSFINETDLHKIEEYCGHSYLEKFNFSQQLWGRKTYLLNKS
FVIOEYFSPNLKFLKFH"
2323..4959
/gene="secA"
/notes="secA"
2323..4959
/gene="secA"
/codon_start=1
/transl_table=11
/protein_id="CAA45961.1"
/db_xref="GI:288998"
/db_xref="SWISS-PROT:Q06461"
/transl_table="MFNFLFSNKINQVIRIVKQINSLSIDSYNNKYSDQELKQTNKLD

CDS
1759..2298
CDS

gene
2323..4959
CDS

QIIATQDIDTILPKAFATKEAIKRAIKRATGLLLFDVLQIGAILNQGKIAEMTKGKTL
VAMLTAYLNSLFNKGVHIVTVNEYLAKRADATLAKQIFEYLNIHIGIDISHSQERKK
OYSCDITLNTSELGFDYLDRDNDMAIOEDLORDFEFAIIDEISLIDISARTPLIIS
GPANNKLTLEYLEANKVANLQNQTDYEIDKKNILNENGIKKSNILDNINLYDIQ
KPWKIYLNAKAKEIFIKNNQYIVARNEIYVDEFTGRIMEGRSSDGLGHAIEAKE
KORIQOENKTLSITONFLYKLSGVTGTAKEAELEIYKLVKVBIPKLNQ
KRLDLSLVKTYEYKWKAVANECFDMYQIGRPTLVGTSTSEKSELAKILKEQVYN
LLNRKPENITRESEITQAGRKYITISTNMAGRTDILGNPQIILAKTALATHINK
ILNLTQYNTNYKIENEITYILNSINNTLNNIDINSODISOSINNTINNMTQDAKS
YKISNIYKVLNKKYKQLCHNEKEIITLGLYVIGTERHESRIDQLSGRSRQDGL
RSSRFLSQDNLKIFGGDKISDFMNLNIDEDMPIESSLNKSSQLQGRSKKEATFY
DVRKQLFEDVLNNORQATYIERKLLKSNYTRDCILEYAESTIEMLVYVYQQQDI
SEKTKILSKLLNINLYNNNLLNMEENDIKSLFEQLRTYIDLEIYSLQLRPG
LIROLEKYLLQOIIDVAMQOEHINKISILKESIGWSYQGDPLIEYKNEAFNLFINV
TYIROTIVYLTMBSRLIVNDN"
BASE COUNT 2239 a 679 c 646 g 1547 t
ORIGIN
Query Match 17.0%; Score 291.4; DB 8; Length 5111;
Best Local Similarity 52.4%; Pred. No. 5e-53;
Matches 765; Conservative 0; Mismatches 676; Indels 18; Gaps 5;
Qy 65 TTATTGATGGATGGGCGCTTTCCGATGAACAACACGGAATGCAATTCCTAAAGCTAAAA 124
Db 240 TTCTTGATGGCTGGGCGCACTCAACTAATGCAAAAGGAAATGCAATTCATGAAGCTAACA 299
Qy 125 CGCCTATTATGGACAAACTTTGTTCTGGAATTTGGCAAAATTTGGA--GCACACGCTC 181
Db 300 CACCAATAATAGACAAACTTTGGGACAAATATATCCAAAACCTTTGTTAAGTGCACTCTCGT 359
Qy 182 TTCATGTTGGATTGCCAGAGGCTTAATCGGAATCTGAATGGAGTGGCAATTTGTAATATAG 241
Db 360 AACATGTTGGCTTACCACATAGCAAAATGGAATTCAGAGTAGGCGCATACACAATCG 419
Qy 242 GAGCTGGAAGAGTTATTTTCAAGATATGTTCGAAATTAATTTGGCTGTTCACAGCAACG 301
Db 420 GGGCAGGAGAAATAAATTAATCAAGACTTAGTAAATAATCAGTAAATATATAAAAAATAAG 479
Qy 302 AGTTTGTACAAATCCTCAGATGTTGCATCAGCTGAGCGTGCAAGAGGGAGTGGTGC 361
Db 480 AATTTTCACAATATTCAATATACATATATATACAGTATTCATGTGAAAAAATTTCTA 539
Qy 362 GATTGCTATTTATGAGACTGGTTAGCGATGGTGGTGGTCCACTCTCATATTTGATCATCTTT 421
Db 540 AAGTACATATTTAGGCTTATGTTCTAATGGTGAGTCCATCTCACATAATCATTTAA 599
Qy 422 TTGGTGTGATGACGTGCAATTTAAACAATACAAATGACGCAAGGTTTTCATTCACATTTT 481
Db 600 AAGCATTAATACAAATTTCAAAACAATATAAAGT---TATGACTTGGCTGCATTTAATTA 656
Qy 482 CTGATGCTCGAGATACCTCCCAACAAGTGGAGCTGGTATCTTTGAACAACCTCTTCAAT 541
Db 657 CTGATGAAGAGATACATTCGCCAAGCAAGCAAAACCTTTTGTGAAGAATAATTATAGATA 716
Qy 542 TTATTGCTCGGAAAAGTACGAGAAATTTGCTACTATTTACTTGGACGTTATTTATGCAATGG 601
Db 717 TAATTGATGAAACTAATAATACGCAATTTCTACAATAAGTGGCAGATATATAGCATGG 776
Qy 602 ATAGGCAAAAGATGGGAGGATTAAGATGGCTTTATAGGCAATTTGTTGGAGGATTTG 661
Db 777 ATAGAGACTGTAGATGGCTAGAAACAGAAAAAGCATACAAGCAATGCTAGAAAAATCAT 836
Qy 662 GACAAAAGCCACGTTTGATAAGGCTCTCGATGTTGTGTAGAGCGATATGCTCAATCTG 721
Db 837 GTAATCAAAACTCTTCTTCTGATATACTATCAATCATCGATAAATATATATGAACAAATA 896
Qy 722 AGACTGACGAATTTCTGAAACCAATTTGTTTTCGGACGATGGCGAGTAAAAAGATGACG 781
Db 897 TCCTGTGATGATTTCTACCACCAACTAAATATCGATGCATAATATTTGT---TCATAACG 953
Qy 782 ATACTCTTATTTCTTCAATTTATCGTCTGATCGTATGCGGCAAAATTTGTAATGTTGG 841
Db 954 ATAGTCTGATTTTTTTTAAATTTTAACTGACAGAAATTAGACAAATTTGCTCCATTCATTTG 1013
```



```
/protein_id="AAK78687.1"
/db_xref="GI:15023590"
/translation="MKFNKKTIEDVDVKGRVLVRCDFNVPLKDGVTIDENRLNGALP
TIKYLSEHGKGVILCSHLGKAGKAGPDSKTLAPVAKRLSELLGREVPAADDTVVGENA
KRAVAELKEGVILLENTFRPEEGNDDAFSKDLASLADVVYVDFAGTAHRAHCSTV
GVTFVDVATCGEYLIOKELKFLGSAGNVPVPVAILGGAKVSKDRIINNLLEKRYDT
LIIGGMATYFLKAQGYTTIGTSLLEADKVDYAKEMIEKAEKGVKLLLPIDNITVGAEF
KADTKATVVEDANIPPEGVGLDIPGKTOKLYADAVEAKTVVWNGPMGVPEFENFAKG
TKDVAKAMAESDATTVIGGDSAAVNQLGFGDKMTHISTGGGASLEFLGKELPGIV
ALNDK"
5411..6157
/gene="CAC0711"
/notes="tpi"
5411..6157
/gene="CAC0711"
/codon_start=1
/transl_table=11
/product="triosephosphate isomerase"
/protein_id="AAK78688.1"
/db_xref="GI:15023591"
/translation="MKPTIAGNKKMNTTISESLKLIBELKPLVKDAKAEVVVAPTAV
SLFTVNAKSGNKKVAAONAHEESGATGEISLKALELGVSYVILGHSERQYFN
ETCALNKKVKAFAFNITPILCCGETLEERANVTNEVTGKQIKLGLAGLSAEQAK
VVIATPEIWAIGTGTATDEQANETIGAIRKTVVEMFGKEVAEKVRIQYGGVYKPNPTI
KAQMKAPEIDGALVGGASLKAADFAAIVNF"
6346..7878
/gene="CAC0712"
/notes="pgm"
6346..7878
/gene="CAC0712"
/codon_start=1
/transl_table=11
/product="2,3-bisphosphoglycerate-independent
phosphoglycerate mutase gene"
/protein_id="AAK78689.1"
/db_xref="GI:15023592"
/translation="MAKPYVLMILDFGDISKVDGNVAKAASKPNFDKYFNPNPTH
LGASVSLGPDQMSNSEVGHNLNIGRIVYQSLTKITKAIEDGDFFKNAALNKAVN
NLVENDSTLHMLGSPGVSHSTNLHGLLOLAKKKNVKVFVHAFLDGRDVPSSA
KEFKIDIEDYNEIGLGETATVSGRYVAMDNDNWERELAYNAMVLGKEEAESAIAK
AVDASHDNKTIDFVLPTVIVKGGKPVATIKDSDVIFNFRPDRAKQITRAIAEAF
DGRKRLNLEFVTMEYDASFGVDVAFGPENITNTLGEYVSNKGLNQLRIAEATKY
ALFNFPDPMVGHFGIILEAKKAVETVDECLGKIVDKVLKLDGSVFTADHGNSEQMDV
YSGNKPMTAHTVNPVPFVYSNHTAEAKKLNIEGLVADIAPTMLQEMGLEKPEENTGKSL
FE"
8000..9295
/gene="CAC0713"
/notes="eno"
8000..9295
/gene="CAC0713"
/codon_start=1
/transl_table=11
/product="Enolase"
/protein_id="AAK78690.1"
/db_xref="GI:15023593"
/translation="MKSVEIVDMARQILDSRANPTVEVVEVLEDGTVGRASVPSGA
STGOFEAELRDNDKQYLGKSVLNAVNDVNETATELIGNVDFDQTLIDQTMLEIDG
TENKSLGANMLGVSLAVARAAYELGISLYQLGVGNKVLVPVMMNVLNGKGHAD
NVDVDFEIMPAGAPSPFSEALRSCAEPVHTLSLQSGLETAAGVDEGGFAPNLNSN
EEAQIIFLEAVTAGYEPKDMFIAMDPASTEYENGKYNLKGEGKYVTSEEMVEVYA
NLVEKYPILISLEDGMAEDWDGKLLTDGRIGDQVQLVGLDGLFVNTKRLSKGLQLVGA
NSLILKLNQIGTITETLANEMAQRAGYTAVSVRSGETEDTTISDLVAVNAGQIKT
GAPARTERVAKYNQLRIEELGEVAEFRGLNAYNIKK"
9420..9656
/gene="CAC0714"
/notes="secG"
9420..9656
/gene="CAC0714"
/codon_start=1
/transl_table=11
/product="Membrane protein secG involved in protein
secretion"
```

```
/protein_id="AAK78691.1"
/db_xref="GI:15023594"
/translation="MRTFLIVAOVLAFALITVLVQPKGMDGFMNSISGCTNETFFAK
NKSKTREAFVLKVLTVLFSVLFAILLIFNLKIFK"
9775..11967
/gene="CAC0715"
/notes="vacB"
9775..11967
/gene="CAC0715"
/codon_start=1
/transl_table=11
/product="FUSION ribonuclease and ribosomal protein S1
domain"
/protein_id="AAK78692.1"
/db_xref="GI:15023595"
/translation="MNIRDNILEFMREQAKPMNIKELSEVFDIGKADYKIFKLLKD
MEKGLIINKRTDHFVPEKMGVLKQGNKAGFGVPEEDRPDVPYIPSNMNGAL
NGDVVAKILKESDSGKKEGIEIRILERANKTVGVYEDRNFQFVADDAIPQDV
FIPKGEKNGAKSGDVVAETVWPKRRNPEGRIVEIGSGKEGVDILTIIKKHLKP
EKPEKVERFAEGIPNEIPESEYKRRDIRDKMVTIDGEDAKDLDDAVSICLSNNGN
FKLGVIHADSVNVRDNPLDKALKRATSVYLDIVPMLPKKLSNGICSLAPRODR
LMSCPMEIDPTGKVIQHEIFESIKTNERMTYTDVTKILLRDHEETIKAFELYDDF
KNMEELASILNKARLLRGAIDDFEESKITLNEUGPEVVKYPERAVANRRIIEFMVLV
CNETIAHFYMANIPFYVRVHEEDPESEKLEFRNEFTHNLGYAVRWGSEHPKQLQDVI
EKIKGKEETVSTLLRLSKQARYSPESCHFGFLAARYCHFTSPTRRYPDLIHRI
MKEYINGRMDKRSKLLIGEDVASHOSSEMERVAQEAEREVDSLKAEMYADMVGBEE
FDGIISSVTNFGIFVOLANTIEGLVHISELTDYVYVDEKYLILVSGERTKKIYKLGDE
ARIKVDKVSVDSEHIEYFKLVENKGERAAQGEIKVEATEVEAEEREEEF"
4724 a 1454 c 2415 g 3469 t
BASE COUNT
ORIGIN
Query Match 16.7% Score 287.2; DB 1; Length 12062;
Best Local Similarity 53.08; Pred. No. 3 5e-52;
Matches 716; Conservative 0; Mismatches 618; Indels 18; Gaps 4;
QY 68 TTGATGATGGGGCCCTTCCGATGACAAACACCGGAATGCAATTCCTAAAGCTAAACGC 127
Db 6377 TAGATGATTTGGAATTTAGACAAAGGTTGATGAAATGCGAATAGAGCTGTAGTAAAC 6436
QY 128 CTATTATGGACAAACTTTGTTCTGGAATTTGGCA---AAAATTGGAAGCACACGGTCTTC 184
Db 6437 CTAATTTGTAATAACTTAATAACTACCCACATACACATCTTTGGGGCAAGCGGATAA 6496
QY 185 ATGTTGATTCGCCAGAGGCTTAATGGAAATCTGAAGTTGGACATTTGAATATAGAG 244
Db 6497 GCGTAGACTTCAGATGGACAATGGGAACTCAGAGTTGGACATTAATAATAGAG 6556
QY 245 CTGGAAGAGTATTATCAAGATATTGTTTGAATTAATTTGGCTTCAACGAAACGAGT 304
Db 6557 CAGGAAGATAGTATTATCAATCGCTTACAAAGATAACAAAGGCAATTTGAAGATGGAGATT 6616
QY 305 TTGTTACAAATCCTCAGATTGTTGCATCAGCTGAGCGTGCNAAAGAGGGAGTGGTCGAT 364
Db 6617 TCTTTAAAAATGACAGCTCTTTAATAAGCTGTTAATAATGCTTTAGAAAATGATTCTACTC 6676
QY 365 TGCATTTATTAGACTCGTTTAGCGATCGTGCTGCCACTCTCATATTTGATCATCTTTTGG 424
Db 6677 TTTCAATTTGGGACTTTTATCTCCAGAGAGGTTCACTCACACACTTAATCATTTAAGG 6736
QY 425 CGTTGATACGTGATTTAAACAATACAAAGTGCACAAAGGTTTTCATTCACCTTTTTCGCT 484
Db 6737 GTCTTTTACAGCTTGCAAGAAGAAGAACGCTAAAAAAGATATTGTTGACATGCTTTTCTTG 6796
QY 485 ATGGTCAGATACTTCCCAACAAGTGGAGCTGGTTATCTTGACACACTTCTTCAATTAA 544
Db 6797 ATGGAAGAGCGTACCCGCTTCTTCTGCAAAAGAAATTTATAAAGATATAGAGACATACA 6856
QY 545 TTGCTTCGAAAAAGTACGAGAAATTTGGCTACTATTACTGGACGTTATTATGCAATGATA 604
Db 6857 TGAATGAATAGGCTTGGTGAATTCGCTACTGTATCTGGAGAGATATTATGCAATGATA 6916
QY 605 GGGACAAAAGATGGGAGCGTATTAAAGATGCTTATGAGGCAATTTGTTGGAGGATTGGAC 664
Db 664
```

Db 6917 GAGATACAGATGGGAAAGAGAAAGAACTTGCCTTACAATGCAATGGTCTTGGAAAAAGGTG 6976
Qy 665 AAAAAGCCCGTTGATAGAGCTGTCGATGCTTTAGAGAGCGATATGCTCAATCTGAGA 724
Db 6977 AGAGGAGAGAGAGCTTATTAAGCTGTAGATGCTTCATATCATGACAAACAAAGATG 7036
Qy 725 CTGACGAATTTCTGAACCAATGTTTTTCGG-----ACGATGGCGGAGTAAAGATG 778
Db 7037 AATTGTGTACTACCAACAGTATAGTTAAAGAGGAAACAGGTTGCGACTATAAGGACA 7096
Qy 779 ACATATCTTATTTCTTCAATATTCGCTGATCGTATGCTCAATTTGTAATGTT 838
Db 7097 AGGATCTGTATATTTCTTCAATTTTTCAGACCTAGACAGCTAGACATACTAGAGCTA 7156
Qy 839 TGGGTCTCGAACCTTATAAGATCTTAATAGTTGCTTCTCACCCTCAAAAATATTGAGA 898
Db 7157 TACAGAGAGAGCTTTGACGGATTTAAGAGAGACAGACTTAATATGAGTTGTAACATA 7216
Qy 899 TTAGTGGGATGACCCCAATACAATTAAGAGTTTCCATTTCCATGTTTATCCCACTGTGA 958
Db 7217 TGACGAATATGATGCAAGCTTTAAGGGAGTTGATGCTGC-----ATTGGACCTGAAA 7270
Qy 959 CTCATCTATGCTGCTGCTGAATGCTTCTTCAAGAGGTTACTCAATTTCACTGTG 1018
Db 7271 ACATCAAAATCTCTTGGAGAATATGTAAGCAACAAAGGTTTAAATCAACTTAGAATTG 7330
Qy 1019 CGAAATGAGAGATCTCTCATGTTACCTTCTTCTTAAATGCTGCTGAGAGTTCAAT 1078
Db 7331 CAGAACTGAAATATGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 7390
Qy 1079 TCCAAGATGAAGAGCTGTTGATGTTTCCGTCACCAAAAGAGTTGCTACATGATTTAA 1138
Db 7391 ATAAAAATGAAGATAGAGCTTTAATATCATCACCA---AGGTTGCAACTTATGATCTTA 7447
Qy 1139 AACCAGAAATGATGCTGCTGGAGTTGCCGAAAAAATGTCGAGCAAAATGAGTCAGGCA 1198
Db 7448 AACCAGAAATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 7507
Qy 1199 GGCATCTTTGTTATGTCGAATTTTGGCCCTCCTGACATGTTGACATGCTGTAAT 1258
Db 7508 AGTATGATGTTTATTAATTAATTTTGTAACTTTGCTAACCTGATGTTGACATGCTG 7567
Qy 1259 TTGAACCTGCGCTCAAGAGCTGCAAGCTGCTGACGAGCAATGTTGAAAGATGTTGAAG 1318
Db 7568 TTGAAGCAGCTAAAGAGCTGTTGAAACAGTTGATGATGATGATGATGATGATGATG 7627
Qy 1319 CATGCCAAATATATGCTTCTTATGTTTCTTCTTCCATGCTGCAATGCTGAGAGCA 1378
Db 7628 AGTTCTTAAATGATGAGAGTGTATTTATACAGCAGATCAGGAAATTCAGAGCAAA 7687
Qy 1379 TGATTGCTCCGATGCTAGTGAACATGCTGCA 1410
Db 7688 TGATCATCTACTATGTAAGAAACCTATGCA 7719

RESULT 15

LOCUS U39725/c 12844 bp DNA linear BCT 19-OCT-2000
DEFINITION Mycoplasma genitalium section 47 of 51 of the complete genome.
ACCESSION U39725 U39730 L43967
VERSION U39725.1 GI:3845017

KEYWORDS
SOURCE Mycoplasma genitalium
ORGANISM Mycoplasma genitalium

REFERENCE 1 (bases 1 to 12844)
Bacteria: Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A.,
Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G.G.,
Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M.,
Fuhrmann, J.L., Nguyen, D.T., Utterback, T., Saudek, D.M.,
Phillips, C.A., Merrick, J.M., Tomb, J., Dougherty, B.A., Bott, K.F.,
Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O. and Venter, J.C.

THE minimal gene complement of Mycoplasma genitalium

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Science 270 (5235), 397-403 (1995)
96026346
7569993

2 (bases 1 to 12844)

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A.,
Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G.G.,
Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V.,
Sandusky, M., Fuhrmann, J.L., Nguyen, D.T., Utterback, T.R.,
Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.,
Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N.,
Smith, H.O., Hutchinson, C.A. III and Venter, J.C.

Direct Submission

Submitted (29-OCT-1995) The Institute for Genomic Research, 9712
Medical Center Drive, Rockville, MD 20850, USA

3 (bases 1 to 12844)

REFERENCE
AUTHORS

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A.,
Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G.G.,
Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M.,
Fuhrmann, J.L., Nguyen, D.T., Utterback, T., Saudek, D.M.,
Phillips, C.A., Merrick, J.M., Tomb, J., Dougherty, B.A., Bott, K.F.,
Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O. and Venter, J.C.

Direct Submission

Submitted (19-OCT-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

On or before Oct 19, 2000 this sequence version replaced

gi:1046149, gi:1046106.

FEATURES

source

1..12844

/organism="Mycoplasma genitalium"

/mol_type="genomic DNA"

/isolate="G37"

/db_xref="taxon:2097"

125..1474

/gene="MG425"

125..1474

/gene="MG425"

/note="similar to GB:U00089 SP:P75172 PID:1673885 percent
identity: 65.56; identified by sequence similarity;

putative"

/codon_start=1

/transl_table=4

/product="ATP-dependent RNA helicase (dead)"

/protein_id="AAC72446.1"

/db_xref="GI:3845018"

/translation="MDSTFHELGISOTLETNALHINKPTKIQIISIPQFLSEKNLI
VHSPGCTGTAFAPLPIELKLEKQAKPTIVIAPTRELVEQIKTFESNIAKKLR
IISLGGVPKQIKKIKTNPOIIVGTGRIMDLERKAIHFSDEHLIIDEVDLMD
RGFKQIFNLLEQINSFKQIAVSYASNOEAINIAKQITNNGIFIGSPFNKADNTNN
DKLIQFQYLFSDQKQALYSLIKTAQVKSIIIVCFDKKGVDDLVHFKKRLTFA
LHGDKQFTRERNLKFANTKQPTILVTVDLIGRGTHVEADIMVHFKKRLTFA
RMGRTRNNHGTCTVCTSOEKKVFLKWEKIKITONRIAECKOMKIKLPLKNKATK
KGGISLDCVQKIYANAKPYDRNKRVPLASDLFPKSRMROPEKAKQKIHNDWQSNM"
1497..1691

gene

/gene="MG426"

1497..1691

/gene="MG426"

/note="similar to GB:U00089 SP:P75171 PID:1673884 percent
identity: 84.38; identified by sequence similarity;

putative"

/codon_start=1

/transl_table=4

/product="ribosomal protein L28 (rpl28)"

/protein_id="AAC72447.1"

/db_xref="GI:3845019"

/translation="MAKKDQLTGLRGLGNRSHSKTITREKWNVNLOSCKIKDTNGK
VTRILVSTKTIKTLKKQNR"

1727..2152

/gene="MG427"

1727..2152

/gene="MG427"

/note="similar to GB:U00089 SP:P75170 PID:1673883 percent
identity: 75.89; identified by sequence similarity;

putative"

Db 5608 ATGCAAGGTAGTATTACATTTTATTTGGTGATGGTAGAGATGTAGCACCTTGTAGCTTAA 5549
 QY 515 CTGGTTATCTTGAACAACTTCTCAATTTATTCCTTCGGAAGAGTACGGAGATTCGGCTA 574
 Db 5548 ACAAAGATCTTGAGAAATTAAGATATTTCTAAAAAATATCCTAATAGTTGTTATTTGGAA 5489
 QY 575 CTATTACTGGACCTTATTTATGCAATGGATAGGCAAAAGATGGGAGCGGTATTAAGATGG 634
 Db 5488 CTATTGGGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5429
 QY 635 CTATTGAGGCAATTTGTTGAGGT-----ATTGGCAAAAAGCCACCGTTGATGATGAGGCTG 688
 Db 5428 CTATAAAGCTTTATTAGGAGTTTCAAAAAAATAATCAATGACCAATTTGGTTATATTG 5369
 QY 689 TCGATCTTCTAGAGAGCGATGCTCAATCTGAGACTGACGAAATTTCTGAAACCAATTG 748
 Db 5368 AAGCGCAATATCAGAACCAAAATTTACTGATGAATTTATTTCTGCAATTAATGCGCAATT 5309
 QY 749 TTTTTCGGACGATGGCGAGTAAAGATGAGGATGACTCTTATTTTCTTCAATTTATCTG 808
 Db 5308 TAAATTTCTGATCAGTTTGCATTAACAATAATGATGAGGTATTTCTTCAATTTATGAG 5249
 QY 809 CTGATCGTATGCTCAAAATTTGTAATGTTGGGCTCGAAGCTTATAAAGATCTTAATA 868
 Db 5248 CTGATAGCAAGACAAATGTCCTATTTGATCTTTAACAGCAATTTATTAACAATCAAC 5189
 QY 869 GTTCGGTCTCCACCCCTAAATAATTTACAGATGATGGGATGACCCCAATCAATAAGACT 928
 Db 5188 CTGAATTTGAACGAAAGAAATTTATTTTGTAAATGATGAATTTATGAGGAAATTG 5129
 QY 929 TTCCATTTCCATCGTTTATTTCCACCTGTGACTACTAATGTGCTTGTGCTGAATGGCTTG 988
 Db 5128 TACCTAGCGAATTTGCTTTTCCACCTCAAAACCATTAATAATAGTCTTGTGAGTAATTG 5069
 QY 989 CTCTCAAGGATTTACTCAATTTCACTGTGCGGAAGTGAAGATGATCTCATGTTACCT 1048
 Db 5068 CTAATAATTAATTTGAGCAATTTGAGGATTTGAGAACTGAGAAATGCTCAGCTTACT 5009
 QY 1049 TCTTCTTTAATGTTGTCGAGAGTTCAATTTCAAGATGAGGCGTTCTATGTTCCGT 1108
 Db 5008 TCTTTTTCATGTTGTTGAGTTAATCTCAGCAATGAAACAGACATTAAATTCCTT 4949
 QY 1109 CACCAAGAAGTTGCTACATATGATTTAAACACAGAAATGATGCTGCTGGAGTTGCCG 1168
 Db 4948 CTTTAA---AAGTGTACATATGATTTAGTCCCGAAATGTCATGTAAGCTATTACTG 4892
 QY 1169 AAAAAATGCTCAGCAAAATTTGACTCAGGAGGATCCTTTGGTTATGTCGAATTTTGGCC 1228
 Db 4891 ATGCACACTAGAAAAGCTTAA-----TAACCTTTGATTTTACTGTTTAAATTTTGC 4838
 QY 1229 CTCTTGACATGTTGGACATGCTGTAATTTGAACCTGCTGCAAGCATGTCAGCTA 1288
 Db 4837 ATCTGATATGTTGATGCTATGTTGATGATGATGATGATGATGATGATGATGATGATG 4778
 QY 1289 CTGACGAGCAATTTGGAAGATTTTGAAGCATGCAAACTTATTAATACGTTCTTATGG 1348
 Db 4777 TCGATGTTCAAAATTAACCAATAGTTGATTTTGTAAAGCTAATCAATAAATGATGTTT 4718
 QY 1349 TTACTTCCGATCATGGAATGCTGAGAGATGATGCTCCCGATGCTGAGTGAACATG 1408
 Db 4717 TAACTGAGATCATGGGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 4658
 QY 1409 CACATACCTGCAATTTGGTCCCATTTACTTCTCTTCAAAACATTTGTTTAAATGCA 1468
 Db 4657 ACACACATTTATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4598
 QY 1469 CT 1470
 Db 4597 CT 4596